NOTE TO USERS

This reproduction is the best copy available.

UMI®

University of Alberta

Molecular Genetics of Host-Pathogen Interactions between *Brassica* species and *Leptosphaeria maculans*

by

Jed Arden Christianson ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in

Plant Biology

Department of Biological Sciences

Edmonton, Alberta
Fall 2004

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-95919-8
Our file Notre référence
ISBN: 0-612-95919-8

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Acknowledgements

First and foremost I would like to thank my parents, Barry and Linda Christianson, without their unflinching support and love I could not have done this (It's finally finished mom). I am grateful to my Supervisor, Dr. Allen Good for all of his guidance and advice and also for providing the opportunity to do this thesis. Thanks to Dr. Derek Lydiate, for accepting me into his lab group and making this project possible. Dr. Andrew Sharpe and Dr. Roger Rimmer each provided input and expertise that were invaluable to this thesis. The bulk of the research presented in this thesis was carried out at the Saskatoon Research Centre of the department of Agriculture and Agri-food Canada; the resources provided to me there are greatly appreciated. For funding and support I am grateful to Bayer Crop Science and especially Dr. Malcolm Devine. I would like to thank Lindsay Yaworsky for her assistance in the construction of the *L. maculans* mutagenized strain library. To Dr. Steve Robinson for all of the illuminating debates, discussions and pints, cheers. Finally, many thanks to all of the other friends and family who have provided me with meals, places to stay and good company through my studies.

TABLE OF CONTENTS

1. General Introduction

1.1. The molecular genetics of host defense........1

1.2. Overview of non-invasive hemibiotrophic pathogen growth *in planta*......4

1.3 The Genus *Brassica*........7

1.3.1. Evolutionary history of the *Brassicas*

1.3.2. The model plant *A. thaliana* as a tool for studying the *Brassicas*

1.3.3. Description of the *Brassica* species discussed in this thesis; *B.napus* and *B.juncea*

1.4. Characteristics and life cycle of *L. maculans*........14

1.4.1 Pathogenicity groups and avirulence genes in *L. maculans*

1.4.2 Molecular genetic studies of *L. maculans* biology

1.5. Molecular interactions in the *Brassica-L. maculans* pathosystem........19

1.6. Research objectives and approaches........20

2. Genetic mapping of genes for resistance to *Leptosphaeria maculans* in *Brassica juncea*

2.1. Introduction........38

2.2. Methods and materials........41

2.2.1. Parental lines and population development.

2.2.2. Preparation of *L. maculans* inoculum

2.2.3. Plant cultivation and disease ratings.

2.2.4. Marker analysis using restriction fragment length polymorphism (RFLP) markers
2.2.5. Microsatellite markers
2.2.6. Genetic mapping
2.3. Results........45
2.3.1 Production and phenotypic analysis of mapping population
2.3.2. Developing a map of the *B. juncea* genome
2.3.3. Mapping the dominant gene for resistance to *L. maculans*, *LMJR1*
2.3.4. Predicting a probable location for a recessive resistance gene to *L. maculans*
2.4. Discussion........55

3. Analysis of EST libraries from *Leptosphaeria maculans* and infected *Brassica napus* leaf tissue

3.1. Introduction........66
3.2. Methods and Materials........69
3.2.1. Culture and growth conditions
3.2.2. RNA extraction
3.2.3. cDNA library construction
3.2.4. Plasmid preparation
3.2.5. cDNA sequencing

3.2.6. Sequence analysis

3.2.7. % G+C content and hexamer frequency analysis

3.3. Results..73

3.3.1. General characteristics

3.3.2. Determining sequence origin in mixed libraries

3.3.3. Analysis of putative functions

3.4. Discussion..89

4. Microarray analysis of genes from *Leptosphaeria maculans* and *Brassica napus* expressed over the course of infection

4.1. Introduction..99

4.2. Methods and materials..102

4.2.1. Plant cultivation, spore production and inoculation

4.2.2. RNA extraction and labeling

4.2.3. Preparation of cDNA microarray slides

4.2.4. Hybridization

4.2.5. Data collection

4.3. Results..105

4.3.1. Microarray hybridization replicates

4.3.2. Identification of genes with significant expression changes

4.3.3. Gene clustering based on expression patterns

4.4. Discussion..119

4.4.1. Consistency of microarray results

4.4.2. Expression of *L. maculans* genes over the course of infection

4.4.3. Expression of *B. napus* genes over the course of infection

4.4.4. Identifying genes with similar expression patterns

4.5. Conclusions..127

5. Identification of putative *L. maculans* knockout mutants from a T-DNA insertion library

5.1. Introduction..134

5.2. Materials and methods..136

5.2.1. Fungal and Bacterial strains

5.2.2. Transformation of *L. maculans* strain WA74

5.2.3. Construction of the transformant library

5.2.4. DNA extraction from pooled transformant spores

5.2.5. Screening the transformant library

5.2.6. Sequencing reactions and analysis

5.2.7. Southern hybridization

5.2.8. Cotyledon inoculations

5.3. Results..144

5.3.1.Construction of the *L. maculans* T-DNA insertion library

5.3.2. Screening the *L. maculans* insertion library

5.3.3. Sequencing T-DNA insertion sites

5.3.4. Phenotypic testing of putative knockouts

5.3.5. Analysis of ILS-409R sequence

5.4. Discussion..156

6. Conclusion and future prospects..169

LIST OF TABLES

Table 2-1. Segregation for resistance in F3 progeny derived from F2 individuals scoring intermediate or susceptible......48

Table 3-1. A list of ESTs of interest from the LM library, presented here are candidates for pathogenicity genes with descriptions of their closest similarity matches......83

Table 3-2. A list of ESTs of interest from the ILS library, presented here are candidates for defense or pathogenicity genes with descriptions of their closest similarity matches......85

Table 4-1. Pearson correlation coefficients (r) for comparisons between replicates......109

Table 4-2. Number of genes determined to be significantly upregulated or Downregulated at the given false discovery rate for each time point comparison according to the Significance Analysis of Microarrays program......111

Table 4-3. Genes exhibiting a significant change in expression......113

Table 5-1. List of ESTs chosen for reverse genetic screening and their closest BLAST search matches......140

Table 5-2. Primer sequences used for reverse genetic screening......141

Table 5-3. Ratings of disease severity caused by putative knockouts on differential *Brassica napus* cultivars......153

LIST OF FIGURES

Figure 1-1. A general schematic of host responses to pathogen infection..............3

Figure 1-2. Genomic relation between diploid and amphidiploid *Brassica* species....9

Figure 1-3. The life cycle of *Leptosphaeria maculans* on *Brassica*.....................15

Figure 2-1. Crossing strategy and population development......................................44

Figure 2-2. Disease ratings for parental genotypes and plants of the segregating F2 and B1 populations...47

Figure 2-3. F2 derived map of the *B. juncea* genome showing alignment of the two subpopulation maps...50

Figure 2-4. Marker scoring data for J1...53

Figure 2-5. Linkage group J13, generated from B1 segregation data.........................56

Figure 3-1. Breakdown of closest related organism for each EST in each library.....76

Figure 3-2. G+C content distribution for sequences from *B. napus* (BN), *L.maculans* (LM) and the ILS library ESTs (ILS)..79

Figure 3-3. G+C content of the ILS library ESTs broken down by organism Classification...80

Figure 3-4. Hexamer composition analysis t-values..81

Figure 4-1. Illustration of the experimental strategy for time point comparisons...107

Figure 4-2. Correlation between Log_2 ratios for different replicate slides of the half day and one day timepoint comparisons...108

Figure 4-3. Gene expression consensus clusters identified by multiple K-medians clustering runs..117

Figure 4-4. Gene expression clusters as identified by hierarchical clustering.........118

Figure 5-1. The Hygromycin resistance T-DNA construct.......................................143

Figure 5-2. Southern analysis of T-DNA insertion events.......................................145

Figure 5-3. Hypothetical insertion structure and available sequence data................150

Figure 5-4. Showing the alignment between the translated *L. maculans* gene represented by the ILS-409 EST and the amino acid sequence of the *C. albicans* Fcr1 (Fluconazole resistance protein 1) gene..155

LIST OF ABBREVIATIONS

ATP	adenosine triposphate
B1	first backcross generation
B1S1	first selfed generation from a first backcross parent
BLAST	Basic Local Alignment Search Tool
bp	base pair
°C	degrees Celsius
cDNA	complementary DNA
Chl	chloroform
cM	centimorgan
CTP	cytidine triphosphate
Cy3; Cy5	Cyanine-3; Cyanine-5
d	day
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
EDTA	ethylene diamine tetra acetic acid
EST	expressed sequence tag
F1	first filial generation
F2	second filial generation
F3	third filial generation
G; mg; µg; ng	gram; milligram; microgram; nanogram
GSS	genomic survey sequence
GTP	guanosine triphosphate

HCl	hydrochloric acid
hrs	hours
IPTG	isopropyl-beta-D-thiogalactopyranoside
JA	jasmonic acid
kb	kilobase
L; ml; µl	litre; milliltre; microlitre
LB	Luria-Bertami
LESTer	*Leptosphaeria* EST Search Tool
LOD	logarithm of the odds
LRR	luecine rich repeat
M; mM; µM	Molar; millimolar; micromolar
Mb	megabase
Min	minute
mm; cm; cm^2	millimetre; centimetre; centimetre squared
mRNA	messenger RNA
Myr	million years
NCBI	National Centre for Biotechnology Information
nt	nucleotide
PCR	Polymerase Chain Reaction
PG	Pathogenicity Group
Phe	phenol
pmol	picomol
PR	Pathogen Response

RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RNA	ribonucleic acid
rRNA	ribosomal RNA
s	second
SA	Salycilic Acid
SAM	Significance Analysis of Microarrays
SDS	sodium dodecyl sulfate
ssDNA	single stranded DNA
t	timepoint
TB	Terrific Broth
TE	tris-EDTA
TIGR	The Institute for Genomic Research
TTP	thymidine triphosphate
χ^2	chi-squared
Xgal	5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside

1. General Introduction

1.1. The molecular genetics of host defense

Genetic approaches to understanding the interactions between plants and their pathogens focus on the question of why one individual or variety of plants is resistant to disease while others are not, and how resistance traits are inherited. Initially, genetic analyses of inherited resistance to plant diseases focused on single or relatively few genes that conferred resistance at a phenotypic level. Gene-for-gene interaction models were developed to explain disease outcomes between different pathogen strains and host cultivars (Flor 1947). Briefly, if a plant host carries a resistance gene that recognizes a specific pathogen avirulence gene no disease occurs. If the plant resistance gene, the pathogen avirulence gene or both are missing, disease develops. It was hypothesized that these resistance genes encoded receptors that specifically recognized and bound pathogen avirulence gene products. Lack of a functioning resistance gene or avirulence gene would render the pathogen invisible to the host and allow disease progression. The cloning of resistance genes did not begin until 1992 when the maize resistance gene Hm1 was isolated (Johal and Briggs1992). This gene product is an exception to the majority of resistance genes as it is responsible for detoxifying the *Cochliobolus carbonum* HC toxin and not for recognizing specific avirulence gene products. Many other resistance genes have been cloned since this first one and nearly all of them share leucine rich repeat (LRR) regions associated with protein-protein binding (Nimchuk et al. 2001). These LRR type resistance genes act as sentries which can identify fungal avirulence factors and trigger a host of metabolic, enzymatic and physiological defense responses. Initially it was thought that resistance genes directly recognized their avirulence cognates. It was not well understood why selection pressures would not modify avirulence gene structure to evade detection, although it was assumed that they must have some function necessary for survival or infection. Evidence for direct binding between resistance and avirulence genes has been rare. Only two pairs of gene products, Pita and AvrPita from the rice-*Magnaporthe grisea* interaction and RRS1-R and PopP2 from the *Arabidopsis thaliana*-*Ralstonia solanacearum* interaction have been shown to interact despite attempts with dozens of other pairs (Jia et al. 2000; Deslandes et al. 2003). In light of this the guard hypothesis was proposed (Dangl and Jones 2001), which holds that avirulence factors are required for pathogenicity and interact with host targets. LRR type resistance gene products monitor these other host targets for conformational changes that indicate pathogen attack. This guard activity has been thoroughly demonstrated for RPS2 in *A. thaliana* which confers resistance to AvrRpt2 carrying strains of *Pseudomonas syringae* by detecting perturbations in RIN4, the *A. thaliana* target of the AvrRpt2 gene product (Mackey et al. 2003).

Whichever way recognition occurs, it is rapidly followed by a series of downstream signaling events culminating in the transcriptional activation of defense response genes and activation of metabolic and physiological deterrents (Figure 1-1). These signaling pathways are complex and overlapping, involving three signaling molecules, salicylic acid, jasmonic acid and ethylene (Kunkel and Brooks 2002). There is significant amount of regulatory cross-talk between these pathways and the degree of overlap and divergence between each signaling pathway is beginning to be elucidated with the aid of techniques

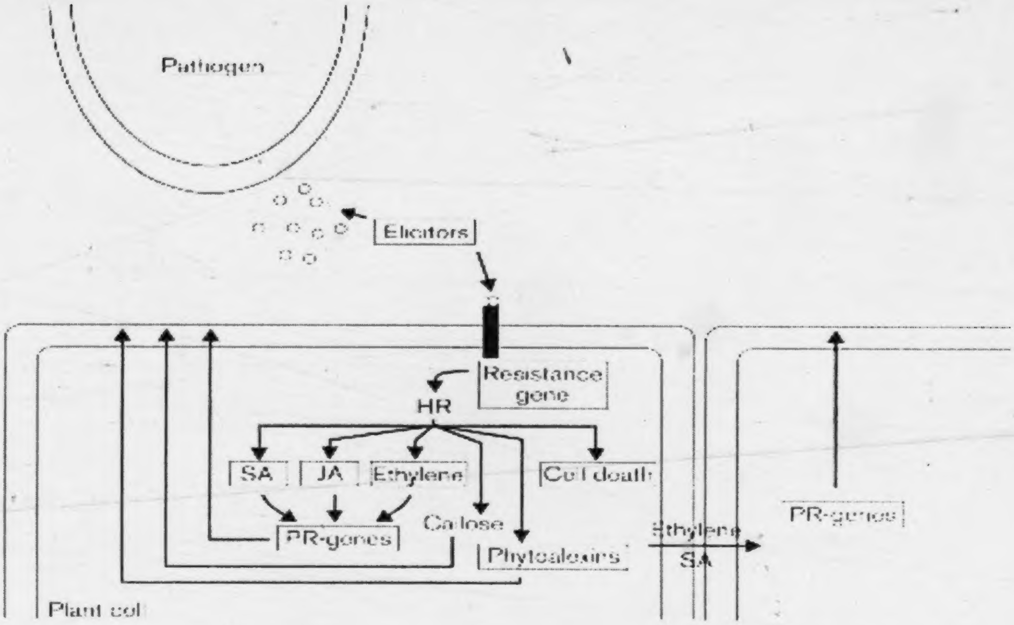

Pathogen
Elicitors
Resistance gene
HR
SA
JA
Ethylene
Cell death
PR-genes
Callose
Phytoalexins
Ethylene
SA
PR-genes
Plant cell

Figure 1-1. A general schematic of host responses to pathogen infection. HR- hypersensitive response; SA- salicylic acid; JA- jasmonic acid; PR- pathogen response. Pathogen elicitors are recognized by host resistance gene products which trigger a hypersensitive response. This response includes the production of hormones involved in induction of pathogen response genes, callose deposition to act as a physical barrier to pathogen growth, release of phytoalexins to directly combat pathogens and triggering programmed cell death to limit pathogen growth. Reproduced from Melchers and Stuiver (2000).

like microarray analysis (Naruska et al. 2003; Schenk et al. 2000). These signaling pathways result in the activation of plant responses to infection involving a wide variety of structural and metabolic changes. Included are defenses such as callose deposition around infected areas, production of active oxygen species and active cell death to block the spread of infection to living tissue (reviewed by Hammond-Kosack and Jones 1996). Although it is possible to sketch a reasonable picture of how plants detect and fight off pathogen attack, not all responses have been well characterized and the relative importance of some of the characterized ones is still unknown.

1.2. Overview of non-invasive hemibiotrophic pathogen growth *in planta*

While much study has gone into the analysis of plant responses to pathogen attack, relatively little attention has been devoted to the pathogens themselves. This is changing, as the number of pathogenicity genes described has been increasing at an exponential rate (Idnurm and Howlett, 2001). These pathogenicity genes are a diverse group broadly defined as genes required for disease development, but not for the completion of a life cycle *in vitro*. They include genes involved in the production of specialized infection structures, enzymatic digestion of plant barriers, defense against plant antifungal compounds, genes required for the production of toxins and of course genes involved in intracellular signaling cascades (Idnurm and Howlett 2001). The diversity of pathogenicity factors reflects the diversity of phytopathogenic fungi and the variety of infection strategies employed by them. Phytopathogenic fungi include facultative or obligate parasites, necrotrophic, and biotrophic or hemibiotrophic pathogens. Each species takes a different approach to invasion, some use specialized infection structures such as appressoria while others seem to utilize less sophisticated tools, relying on wounds or natural openings for entry. Once inside a host, various fungi employ different methods for obtaining nutrition. Some actively kill and enzymatically digest dead tissue, some siphon nutrients from live host cells using dedicated feeding structures such as haustoria, and some grow without damaging or invading host cells, seeming to scrounge what they can from the apoplast.

In the case of biotrophic and hemibiotrophic pathogens such as *Leptosphaeria maculans* that do not necessarily penetrate or kill host cells, there are questions concerning the nutrition and survival of the pathogen in the host apoplast. Research on *Cladosporium fulvum,* a pathogen of tomato, has been conducted concerning the nutritional limiting factors to these fungal hemibiotrophs. Conditions of carbon and nitrogen starvation induce a number of genes which are also induced *in planta* (Coleman et al. 1997). Expression of the *Avr9* gene of *Cladosporium fulvum* is increased in response to nitrogen starvation and is regulated by the nitrogen response factor *Nrf*1, although *Nrf*1 knockouts are still avirulent on *Cf*-9 containing tomato plants (Perez-Garcia et al. 2001). Methionine synthase is induced during infection of tomato by *Cladosporium fulvum* although it is repressed *in vitro* when exogenous methionine is supplied, indicating a possible lack of methionine available in the plant apoplast (Solomon et al. 2000). It was later demonstrated that a number of amino acids are available in limited quantities during most of the infection, although in compatible infections the amounts of certain amino acids, including methionine and γ-aminobutyric acid increases in the apoplast (Solomon and Oliver 2001). Methionine synthase is expressed at high levels at the same time that methionine availability increases in the apoplast, however despite the increase, overall the concentration of methionine remains very low and may not be enough to repress methionine synthase expression. Interestingly, it was shown that *Gat*1, a *C. fulvum* gene required for γ-aminobutyric acid metabolism is expressed at high levels, at the same time that a tomato gene responsible for the synthesis of γ-aminobutyric acid is also highly expressed in a compatible infection, although not in an incompatible one (Solomon and Oliver 2002). This would suggest that *C. fulvum* is capable of spurring the production of γ-aminobutyric acid in tomato to provide a nitrogen source.

Aside from obtaining resources to grow and develop, biotrophic pathogens must also evade detection during invasion of a host. A proline rich glycoprotein named *CIH*1 has been cloned from the hemibiotrophic fungus *Colletotrichum lindemuthianum* (Perfect et al. 1998). This protein is expressed *in planta* but not in culture, and is localized to the fungal cell wall; it was proposed that the protein may be involved in avoidance of recognition by the plant.

There is still much to be learned about the survival of biotrophic and hemi-biotrophic pathogens in host environments. The molecular weapons available to a fungus during colonization are largely uncharacterized, and any pathways leading from recognition of a host and signaling this recognition are almost wholly unknown. While general themes may emerge, it is certain that each specific host-pathogen interaction will have its own peculiarities, given the diversity of tactics and approaches already recognized in plant pathogenic fungi.

1.3 The Genus *Brassica*

The brassicas are dicotyledonous plants characterized by branching racemes, early leaves which tend to be lobe shaped and attached to the stem in an alternate fashion by a stalk, and later leaves which are longer, narrow, tapering and have a saggitate base wrapping around the stem. *Brassica* species produce characteristic flowers ranging in color from white to yellow with four petals arranged in a cruciform fashion, four sepals and six anther bearing stamens. Fertilized flowers give rise to long, slender seed bearing siliques. *Brassica* species include a diverse array of vegetable, condiment and oilseed crops including broccoli, radishes, turnips, mustard, oilseed rape and canola, making this group very economically important. Of greatest importance in Canada are the species *Brassica napus*, *Brassica rapa,* and *Brassica juncea* as sources of edible canola oil, and *Brassica oleracea* as a source of cole vegetable crops. The related genera *Raphanus*, which includes the common radish and *Sinapis*, which is a source of condiment mustard along with *Brassica nigra*, are also cultivated in Canada. Some races of *Brassica* and related genera have also become naturalized weeds such as wild mustard and wild rape. The model plant *A. thaliana* is also a close relative of the *Brassica* species. The combination of economic importance and close relation to a well studied model plant makes the crop *Brassica* species attractive to study. Blackleg caused by *L. maculans* infection is a major disease of *Brassicas,* causing heavy yield losses in canola crops worldwide. In the interests of increasing canola production, many research efforts are devoted to understanding the mechanisms of this disease and host resistance to it.

1.3.1. Evolutionary history of the genus *Brassica*

Despite having very different morphologies, the brassicas are very closely related. Species from five different genera can be divided into two major lineage groups, the *B. oleracea / B. rapa* group and the *B. nigra* group based on RFLP analysis of the chloroplast genome (Warwick et al. 1992). The presence of species from the *Sinapis, Diplotaxis* and *Brassica* genera in both groups indicates that original classifications based on morphological characters do not correspond to their phylogeny. Like many plants, the brassicas have large amounts of duplication within their genomes. There is evidence of multiple historic polyploidy events in the evolution of this group and many current member species are polyploids. The complete nucleotide sequence of *A. thaliana* indicates that this species, with a genome size of only 125 Mb (compared to ~650 Mb for *B. oleracea*) is the result of an historical tetraploidization event occurring ~112 Myr ago (The *Arabidopsis* Initiative 2000). This tetraploidization occurred considerably before the estimated divergence of *Arabidopsis* and *Brassica* between 14.5 and 20.4 Myr ago (Yang et al. 1999). Subsequent to this tetraploidization event, a further triplication appears to have occurred in the *Brassica* lineage (Parkin et al. 2002).

The modern *Brassica* species can themselves be divided into two categories based on their levels of polyploidization. Of the six major *Brassica* species, three are diploid; *B. rapa* (AA), *B. nigra* (BB) and *B. oleracea* (CC). The other three are amhidiploids, comprised of different pairings of the diploid genomes; *B. juncea* (AABB), *B. napus* (AACC) and *B. carinata* (BBCC) (U 1935). The component genomes in the amphidiploid species are very well conserved with respect to their diploid ancestors (Figure 1-2). A resynthesized *B. napus*, created from an interspecific *B. rapa* X *B. oleracea* cross, and a

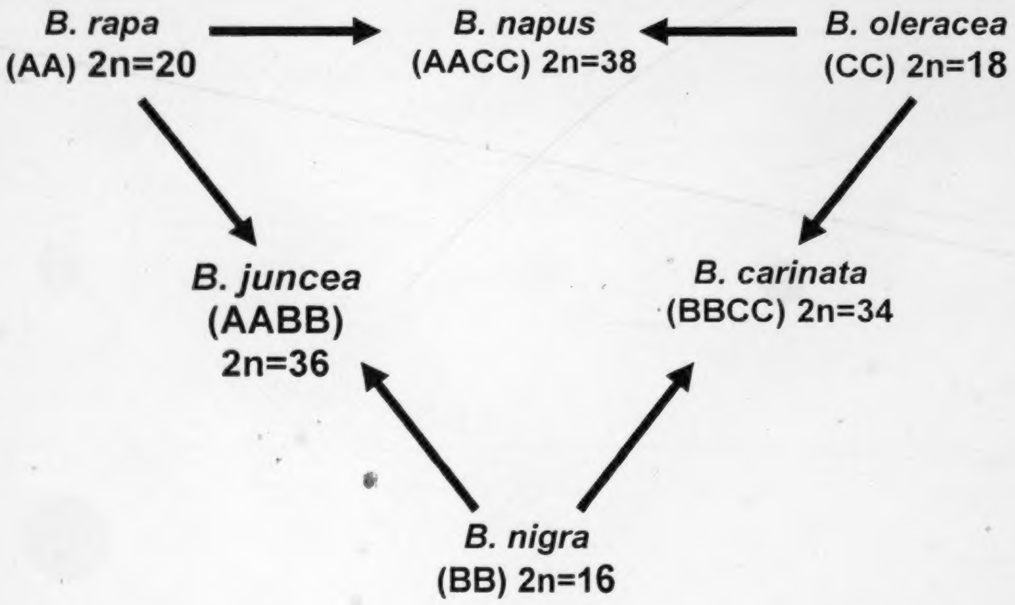

B. rapa
(AA) 2n=20
B. napus
(AACC) 2n=38
B. oleracea
(CC) 2n=18
B. juncea
(AABB)
2n=36
B. carinata
(BBCC) 2n=34
B. nigra
(BB) 2n=16

Figure 1-2. Genomic relation between diploid and amphidiploid *Brassica* species, redrawn from U (1935).

natural *B. napus* accession were used as parents to create a mapping population (Parkin et al. 1995). An examination of restriction fragment length polymorphism (RFLP) markers in this population revealed that their inheritance was predominantly disomic. This indicated that the *B. rapa* derived chromosomes in the resynthesized parent were pairing exclusively with identifiable A genome homologues in *B. napus* and similarly the *B. oleracea* derived chromosomes were pairing with C genome homologues. A genetic map of the *B. oleracea* genome was constructed using the same set of RFLP markers (Bohoun et al. 1998). Strong collinearity was observed between the two C genome maps. This common set of RFLP markers has also been applied to the genomes of *B. juncea* (Axelsson et al. 2000) and *B. nigra* (Lagercrantz and Lydiate, 1995). In the case of the *B. juncea* map, it was possible to identify the A and B genome linkage groups and align them with their homologues from *B. nigra* and *B. napus* (Axelsson et al. 2000). Marker based examinations of the ancestral A, B and C, genomes have demonstrated that they each contain eight triplicated regions covering near the entirety of their genomes (Lagercrantz and Lydiate 1996). Comparisons between these three genomes revealed that although there were major chromosomal rearrangements between them, regions of collinearity covering the entirety of each of the genomes were present.

1.3.2. The model plant *A. thaliana* as a tool for studying the *Brassicas*

A. thaliana has long been a model plant due to its small size, rapid generation times, small genome and ease of transformation with *Agrobacterium tumefaciens*. The complete *A. thaliana* genome has been sequenced, the first plant for which this was accomplished (The *Arabidopsis* Initiative 2000). Numerous annotated datasets which include sequence polymorphisms including alleles, germplasm stocks and phenotypes, gene ontology annotations, gene families, protein information, metabolic pathways, gene expression data from microarray experiments and seed and DNA stocks are now available for this organism (Rhee et al. 2003). Microarrays containing extensive *A. thaliana* gene sets are available from a number of commercial suppliers. Large populations of Arabidopsis T-DNA insertion and fast neutron mutagenized lines have been developed for reverse genetic screens (Krysan et al. 1999; Sessions et al. 2002; Li and Zhang 2002).

Large segments of marker collinearity have been demonstrated between *A. thaliana* and *B. napus* (Cavell et al. 1998; Parkin et al. 2002). Coding region sequences have been shown to be 87% similar between *A. thaliana* and *B. napus* (Cavell et al. 1998). Given the close likenesses between these two species at a genetic level, it is very easy to relate findings in one species to the other. Genomic and EST sequences available for *A. thaliana* in particular can be useful in characterizing sequences from *Brassica* species. *A. thaliana* EST sequences with significant homology to resistance genes were used as molecular markers to identify possible resistance gene candidates in *B. napus* where few resistance gene loci had previously been characterized (Sillito et al. 2000). In light of difficulties presented by aberrant meiotic behavior, genomic sequence for *A. thaliana* was used to more fully characterize a *Raphanus sativa* introgression carrying the *ogura* restorer gene in *B. napus* and to identify an *A. thaliana* genomic region containing the ogura restorer ortholog (Giancola et al. 2003). Comparisons between the genomes of *Capsella rubella* (another close relative of *A. thaliana* and *Brassica* species) and *A. thaliana* were greatly aided by the availability of annotated genomic sequence data for this species. Examination of evolutionary sequence conservation between the two species in coding and non-coding regions demonstrated that sequence divergence was greater in non-coding regions (Boivin et al. 2004). Microarrays constructed from *A. thaliana* sequences have also been demonstrated to be useful for interrogating transcript abundance in *B. napus* mRNA samples (Girke et al. 2000). *Brassica* researchers can make use of readily available, relatively low cost commercially produced microarrays containing various sets of *A. thaliana* sequences. Unfortunately despite being a close relative of the *Brassicas*, *A. thaliana* is a non-host to *L. maculans*. Studies of the interaction between *Brassica* and *L. maculans* can only take advantage of the genomic tools developed from *A. thaliana* indirectly.

1.3.3. Description of the *Brassica* species discussed in this thesis; *B. napus* and *B. juncea*

B. napus is an amphidiploid *Brassica* species containing the A and C genomes. It has a haploid chromosome number of 19. This species is cultivated as the vegetable crop rutabaga, but its primary value is in the production of oilseed. Production of *B. napus* oilseed rape increased dramatically following the development of canola varieties. The term canola refers to any oilseed *Brassica* that contains less than 2% erucic acid in the oil and fewer than 3 mg/g of glucosinolates in the crushed meal. *B. napus* is the most widely grown canola species, although many *B. rapa* varieties meeting the criteria are also available and currently four varieties of canola quality *B. juncea* are registered with the Canadian Food and Inspection Agency (Government of Canada). The great majority of *B. napus* canola varieties grown in Canada are spring types although winter types exist and are commonly cultivated in other parts of the world.

The majority, 65.7% of 135 natural *B. napus* spring type accesions tested, were completely susceptible to infection by *L. maculans* (Rouxel et al. 2003). Many breeding efforts have been directed at developing resistance to this disease. To aid breeding efforts and better understand the genetic nature of some resistance traits better, some sources of *B. napus* resistance have been characterized in more detail. The resistance trait *LmFr1*, derived from the French cultivar Cresor was localized on a genomic map constructed from RFLP markers (Dion et al. 1995). A resistance trait from the cultivar Major, *LEM1* was identified and genetically mapped (Ferreria et al. 1995). Only one other report of a genetically mapped *B. napus* resistance trait has been published, describing *LmR1* from the Australian cultivar Maluka (Mayerhofer et al. 1997).

B. juncea is another amphidiploid *Brassica* species containing the A and B genomes. It has a haploid chromosome number of 18. Canola quality oilseed varieties of *B. juncea* have recently been bred (Cheng et al. 2001). However, the majority of *B. juncea* is still grown for condiment mustard. The development of *B. juncea* as an edible oilseed crop has been spurred by its agronomic advantages such as heat and drought tolerance, reduced seed shattering and resistance to blackleg (Woods et al. 1991). B genome containing *Brassica* species have very high, though not complete levels of resistance to blackleg (Rimmer and van den Berg 1992; Purwantara et al. 1999). Various sources of *B. juncea* resistance have been examined (Pang and Halloran 1996; Plieske et al., 1998; Dixelius, 1999). All of these studies involved introgressions of B genome material into a *B. napus* background due to the rarity of susceptible *B. juncea* cultivars and because the ultimate goal of working with B-genome resistance is to introduce it stably into related genera.

1.4. Characteristics and life cycle of *L. maculans*

The filamentous ascomycete *L. maculans* (also known as *Phoma lingam*, the name given to the imperfect state of the fungus), is the causal agent of blackleg disease on brassicas. Blackleg is one of the most important diseases of *Brassica*, causing damage to crops worldwide. Due to its impact on canola production there is great interest in studying the biology of *L.* maculans, particularly factors that contribute to infection and pathogenicity.

The fungus *L. maculans* grows septate, branched, hyaline mycelia when young, but these turn darker with the buildup of pigment over time. This fungus produces wind borne ascospores after mating, which are thought to be the primary source of infection (Howlett et al. 2001). Asexual pycnidiospores may also be produced; these emerge in pinkish exudates and are likely spread to neighboring plants through rain splash (Petrie 1978; Howlett et al. 2001). *L. maculans* can survive saprophytically on stubble and crop residue and may persist in field conditions for a few seasons in this manner. The life cycle of *L. maculans* is presented in Figure 1-3. After spores find themselves on a plant, fungal hyphae grow randomly across the surface and enter through wounds or natural openings such as stomata (Hammond and Lewis 1987a; Chen and Howlett 1996). After entering the plant, the hyphae continue to grow, establishing small

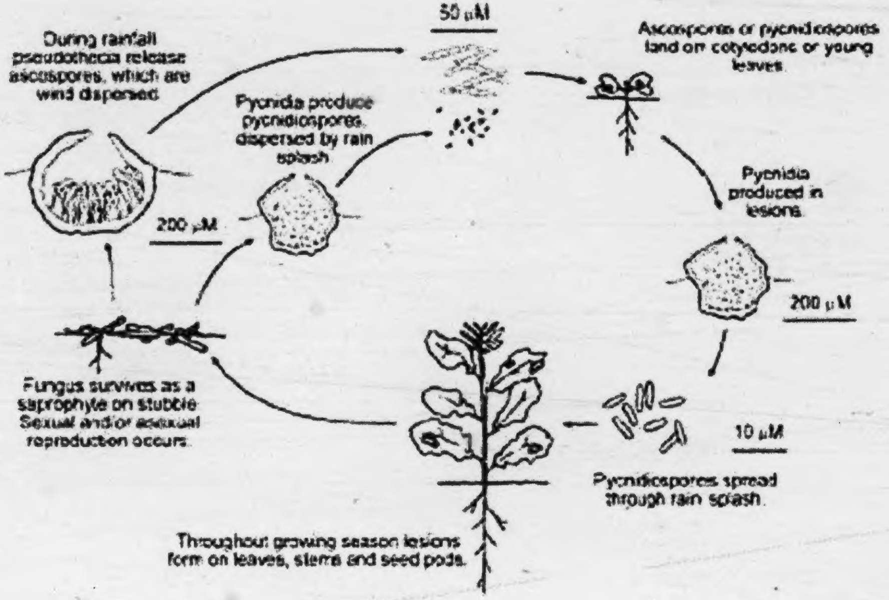

50 μM
During rainfall pseudothecia release ascospores, which are wind dispersed
Ascospores or pycnidiospores land on cotyledons or young leaves
Pycnidia produce pycnidiospores, dispersed by rain splash
200 μM
Pycnidia produced in lesions
200 μM
Fungus survives as a saprophyte on stubble. Sexual and/or asexual reproduction occurs.
10 μM
Pycnidiospores spread through rain splash.
Throughout growing season lesions form on leaves, stems and seed pods.

Figure 1-3. The life cycle of *Leptosphaeria maculans* on *Brassica*. (Howlett et al. 2001)

colonies within the mesophyll by four days post infection (Chen and Howlett 1996). Growth continues biotrophically in the intercellular space until approximately eight to ten days post infection. During this period no symptoms are visible. Around eight to ten days after infection chlorotic patches and necrotic lesions begin to develop on infected plant tissues. Asexual reproduction, involving the production of fruiting bodies called pycnidia, often occurs within these necrotic lesions. These symptoms develop behind the growing hyphal front, which invades the petiole and proceeds to colonize the vascular system (Hammond and Lewis 1987a; Hammond and Lewis 1987b). Growth proceeds in a biotrophic, symptomless manner until the hyphae reach the base of the stem. At this point they invade and kill host cells, giving rise to the characteristic stem canker from which the disease name is derived (Hammond and Lewis 1987b). This canker phase, which inhibits the transport of water and nutrients to the upper parts of the plant, is responsible for the yield loss associated with blackleg. In severe cases plants may become lodged and die before setting seed.

1.4.1 Pathogenicity groups and avirulence genes in *L. maculans*

The various strains of *L. maculans* can be divided into groups based on their virulence towards different varieties of *B. napus* which express different levels of resistance. Initially *L. maculans* was divided into highly and weakly virulent categories, also known as A and B groups. These groups differ not only in their virulence on various *Brassica* species, but also in the production of characteristic metabolites, karyotype, and growth in culture (Pedras et al. 1996; Morales et al. 1993). Although different strains could be successfully crossed within each group, no successful crosses could be made between the two groups (Somda et al. 1997). Subsequently, the weakly virulent group B strains have been reclassified as a separate species, *Leptosphaeria biglobosa* n.sp. (Shoemaker and Brun 2001). Within the A group, or *L. maculans* proper, fungal strains are normally assigned to one of three pathogenicity groups; PG2, PG3 and PG4 (PG1 was originally reserved for weakly virulent B group isolates which are now considered to be a different species). Assignment is based on a particular strain's ability to infect three differential *B. napus* cultivars; Westar, Quinta and Glacier (Koch et al. 1991). Each of these pathogenicity groups can be (and often have been) subdivided into smaller subgroups (Badaway et al. 1991; Kuswinanti et al. 1995; Keri 1999). Genetic analyses have begun to narrow host/pathogen specificities down to gene-for-gene relationships between resistance genes and their avirulence cognates. The first avirulence gene identified was *AvrLm1*, which conditions virulence/avirulence on the differential cultivar Quinta (Ansan-Melayah et al. 1995). This was followed by the characterization of *AvrLm2*, *alm1* and *AvrLm4* which confer avirulence on the *B. napus* cultivars Glacier, Major and Jet Neuf respectively (Ansan-Melayah et al. 1998; Pongam et al. 1998; Balesdent et al. 2001).

A comprehensive analysis of interactions between different pathogen strain/host plant pairings confirmed the presence of four previously identified avirulence genes involved in gene-for-gene interactions and detected the existence of four others (Balesdent et al. 2002). Although characterized genetically, to date none of these genes have been cloned or studied molecularly.

1.4.2 Molecular genetic studies of *L. maculans* biology

To date there has been relatively little work done to study the molecular genetics of *L. maculans*; most published reports focus on electrophoretic karyotyping and genetic mapping. Two independent maps of the *L. maculans* genome have been constructed using different sets of amplified fragment length polymorphism (AFLP) markers (Pongam et al 1998; Cozijnsen et al. 2000). *L. maculans* appears to have quite a plastic genome. Karyotyping of *L. maculans* genomes from various strains has identified variation in chromosome number and size (Taylor et al. 1991). Further karyotyping demonstrated that chromosomal length polymorphisms, possibly as a result of recombination between homologues that can vary up to 35% in length due to the presence of repeat elements, can occur after meiosis (Plummer and Howlett 1993; 1995). Aside from large scale chromosomal polymorphisms, dispensable chromosomes which are inherited in a non-Mendelian fashion and small linear plasmids have also been observed (Leclair et al. 1996; Lim and Howlett 1994).

The first gene from *L. maculans* to be cloned and characterized was nitrate reductase (Williams et al. 1994). Since then only 27 entries containing full gene coding sequences and a set of 121 ESTs have also been submitted to public databases (**http://www.ncbi.nlm.nih.gov/;** Howlett et al. 2001). Some of the cloned genes were selected for study based on the possibility of being involved in infection. Among these are cyanide hydratase (cht), two cellulases (cel1 and cel2), and an endopolygalacturonase (pg1), (Sexton and Howlett 2000; Sexton et al. 2001). The cyanide hydratase and cellulase 2 genes have been demonstrated to be expressed during infection, but no further characterization of the potential role in pathogenesis has been achieved. One gene involved in the glyoxylate cycle, isocitrate lyase was disrupted by the restriction enzyme-mediated integration of a plasmid bearing the hygromycin resistance gene and demonstrated to be required for pathogenicity on *B. napus* (Idnurm and Howlett 2002). Another loss of pathogenicity disruption mutant had also been identified, however the sequence for this particular gene had no matches to any previously identified sequences (Idnurm and Howlett 2003). Recently, methods for *Agrobacterium tumefaciens* mediated transformation of *L. maculans* and selection of homology based targeted disruptions of genes have been established (Gardiner and Howlett 2004).

1.5. Molecular interactions in the *Brassica-L. maculans* pathosystem

To date, research on the plant side of the interaction between various *Brassica* species and the fungal pathogen *L. maculans* has focused exclusively on the identification and genetic analysis of heritable traits for resistance. Although a few of these have been successfully mapped within *Brassica* genomes, none have been cloned and their detailed functions are unknown. Although a general picture of plant defense responses has emerged, it cannot address specific questions about *Brassica* responses to *L. maculans*. Even in *Brassica* hosts which are resistant, the pathogen appears to be able to survive biotrophically and evade detection for a period. Why does resistance derived from the *Brassica* B genome appear to be more complete? Why do resistance responses and disease symptoms appear after a period of silent pathogen invasion? Is there a change in *L. maculans* at this time that de-cloaks it to the host defenses? Do *Brassica* resistance genes function as guards, sensing attacks on host targets after the pathogen begins necrotrophic growth?

Research on the fungal side of the interaction has similarly focused on the genetic characterization of avirulence genes. To date eight of these have been identified, but none has been cloned or studied in molecular detail (Balesdent et al. 2002). Only two pathogenicity genes have been discovered for *L. maculans* and only one of these has a known function (Idnurm and Howlett 2002; Idnurm and Howlett 2003). In contrast to plants, where defense pathways and resistance genes have been conserved, avirulence genes exhibit wide variety of different types. Consequently there is even less understanding of what may be occurring at a molecular level within *L. maculans* during infection. There are many questions about how the fungus survives during its biotrophic phase. How does it receive required nutrients? How does it evade detection during its biotrophic growth phase? What initiates the shift from biotrophic to necrotrophic growth? How does it determine which way to grow to reach the base of the stem?

1.6. Research objectives and approaches

The goal of the research described in this thesis was to develop a suite of complementary molecular genetic tools for the study of *Brassica* resistance to blackleg and *L. maculans* biology, focusing on pathogenicity. Four different genomic scale resources were developed to study genomic location of resistance traits, gene sequence, gene expression and finally gene function in either or both *L. maculans* and its *Brassica* host.

To study *Brassica* resistance to blackleg, natural *B. juncea* populations segregating for resistance to a PG2 isolate of *L. maculans* were developed. Much recent research concerning blackleg resistance has focused on B genome containing species due to their high levels of resistance (Pang and Halloran 1996; Chevre et al. 1997; Keri et al. 1997; Dixelius 1999; Dixelius and Wahlberg 1999; Plieske and Struss 2001; Wretblad et al. 2003). All of these studies suffer the disadvantage of using plant materials containing interspecific introgressions which do not recombine readily with homeologous regions of their host genomes. This results in the appearance of close linkage between loci on the introgressed segment despite potentially very large physical distances, limiting the utility of molecular markers for applications such as positional cloning or breeding. The natural populations described here were designed to enable the genetic mapping of B genome genes for resistance to *L. maculans* with higher resolution between markers than has previously been achieved. Molecular markers flanking these genes can be used to facilitate future efforts to clone them. One gene for resistance, *LMJR1* has been successfully located on a genetic map of the *B. juncea* genome. A second, recessive, gene for resistance has also been demonstrated to be segregating in these populations. However further research is required to identify a precise location for this trait.

Although resistance and avirulence genes are vitally important for determining disease outcomes, the molecular processes of infection and defense are much more complicated. Very little is known about the molecular interactions between *Brassica* species and *L. maculans*. Libraries of expressed sequence tags (ESTs) were created from *L. maculans* tissue and infected *B. napus* leaves as a resource for gene discovery. These EST libraries were constructed to gain more sequence data from *L. maculans* as well as to enrich for and isolate genes expressed by both host and pathogen during an infection. The sequence data gathered is vital to understanding the molecular biology *L. maculans* and its interaction with *Brassica* hosts. These sequences provide the foundation for experimental approaches aimed at determining the biological role of particular genes. A total of 2131 EST sequences were collected and analyzed, 741 from *in vitro* cultured *L. maculans* and approximately another 140 sequences from the infected leaf library are fungal in origin.

Collections of ESTs represent a static snapshot of gene expression at the time that a particular tissue sample is taken. To gain a dynamic picture of the changes in expression of these genes over the course of infection, glass slide cDNA microarrays spotted with all 2131 EST sequences were constructed. Dramatic increase or decrease in expression of a transcript under a particular condition (in this case *L. maculans* infection) implies a connection between the condition and that transcript, aiding in the discovery of genes important for responses to said condition. Gene expression patterns may also aid in assigning putative functions to uncharacterized sequences. These microarrays were constructed for the purposes of identifying genes from both *B. napus* and *L. maculans* which may play important roles in the interaction, including those with no as yet postulated function. A set of 76 genes were identified which had significant expression changes (upregulated or downregulated) between at least two time points after infection. Clusters of genes exhibiting similar expression patterns were identified. Potential roles for these genes in the plant-pathogen interaction are discussed.

Finally, in order to study *L. maculans* gene function in detail, a library of T-DNA tagged insertion mutants suitable for forward and genetic screens was created. Studying the phenotypic effects resulting from the functional loss of a gene of interest is an important key to understanding the biological role that gene plays. To demonstrate the utility of an insertion library for functional genomics and to analyze genes with a potential role in pathogenesis, a reverse genetic screen was conducted to search for T-DNA insertions near a set of 28 candidate genes. Four putative loss of function candidates were found and tested for changes in pathogenicity.

Works Cited

Ansan-Melayah D., Balesdent M.H., Buee M., Rouxel T. (1995) Genetic characterization of *AvrLm1*, the first avirulence gene of *Leptosphaeria maculans*. Phytopathology 85: 1525-1529.

Ansan-Melayah D., Balesdent M.-H., Delourme R., Pilet M. L., Tanguy X., Renard M., and Rouxel T. (1998) Genes for race-specific resistance against blackleg disease in *Brassica napus* L. Plant Breed. 117: 373-378.

Axelsson T., Bowman C.M., Sharpe A.G., Lydiate D.J., Lagercrantz U. (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43: 679-688.

Badawy H. M. A., Hoppe H. H., Koch E. (1991) Differential reactions between the genus *Brassica* and aggressive single spore isolates of *Leptosphaeria maculans*. J. Phytopathol. 131: 109-119.

Balesdent M. H., Attard A., Ansan-Melayah D., Delourme R., Renard M., Rouxel T. (2001) Genetic control and host range of avirulence toward *Brassica napus* cultivars Quinta and Jet Neuf in *Leptosphaeria maculans*. Phytopathology 91: 70-76.

Balesdent M. H., Attard A., Kühn M. L., Rouxel T. (2002) New avirulence genes in the phytopathogenic fungus *Leptosphaeria maculans*. Phytopathol. 92: 1122-1133.

Bohuon E.J., , Keith D.J., Parkin I.A.P., Sharpe A.G., Lydiate D.J. (1998) Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor. Appl. Genet. 93: 833-839.

Boivin K., Acarkan A., Mbulu R.S., Clarenz O., Schmidt R. (1998) The Arabidopsis genome sequence as a tool for genome analysis in Brassicaceae. A comparison of the Arabidopsis and *Capsella rubella* Genomes. Plant Physiol. 2004. 135: 735-744.

Cavell A.C., Lydiate D.J., Parkin I.A.P., Dean C., Trick M. Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome 41: 62-69.

Chen,C.Y., Howlett B.J.(1996) Rapid necrosis of guard cells is associated with the arrest of fungal growth in leaves of Indian mustard (*Brassica juncea*) inoculated with avirulent isolates of *Leptosphaeria maculans*. Physiol Mol Plant Pathol 48: 73-81.

Cheng B.F., Seguin-Swartz G., Somers D.J., Rakow G. (2001) Low glucosinolate *Brassica juncea* breeding line revealed to be nullisomic. Genome. 44: 738-741.

Chevre A.M., Barret P., Eber F., Dupuy P., Brun H., Tanguy X., Renard M. (1997) Selection of stable *Brassica napus-B. juncea* recombinant lines resistant to blackleg (*Leptosphaeria maculans*). 1. Identification of molecular markers, chromosomal and genomic origin of the introgression. Theor Appl Genet 95: 1104-1111.

Coleman M., Henricot B.; Arnau J., Oliver R.P. (1997) Starvation-induced genes of the tomato pathogen *Cladosporium fulvum* are also induced during growth in planta. Mol. Plant Microb. Interact. 10:1106-1109.

Cozjinsen A. J., Popa K. M., Roll B. D., Purwantara A., Howlett B. J. (2000) Genome analysis of the plant pathogenic fungus *Leptosphaeria maculans*: Mapping mating type and host specificity locus. Mol. Plant Pathol. 1: 293-302.

Dangl JL, Jones JD. (2001) Plant pathogens and integrated defence responses to infection. Nature 411: 826-833.

Deslandes L., Olivier J., Peeters N., Feng D.X., Khounlotham M., Boucher C.. Somssich I., Genin S., Marco Y. (2003) Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proc Natl Acad Sci USA 13: 8024–8029.

Dion Y., Gugel R.K., Rakow G.F.W., Seguin-Swartz G., Landry B.S. (1995) RFLP mapping of resistance to the blackleg disease [causal agent, *Leptosphaeria maculans* (Desm.) Ces. et de Not.] in canola (Brassica napus L.). Theor. Appl. Genet. 91: 1190-1194.

Dixelius C. (1999) Inheritance of the resistance to *Leptosphaeria maculans* of *Brassica nigra* and *B. juncea* in near-isogenic lines of *B. napus*. Plant Breeding 118: 151-156.

Dixelius C., Wahlberg S. (1999) Resistance to *Leptosphaeria maculans* is conserved in a specific region of the *Brassica* B-genome. Theor Appl Genet 99: 368-372.

Ferreria M.E., Rimmer S.R., Williams P.H., Osborn T.C. (1995) Mapping Loci controlling *Brassica napus* resistance to *Leptosphaeria maculans* under different screening conditions. Phytopathology, 85: 213-217.

Flor H.H. (1947) Host-parasite interactions in flax rust: its genetics and other implications. Phytopathology 45: 680–685.

Gardiner D.M., Howlett B.J. (2004) Negative selection using thymidine kinase increases the efficiency of recovery of transformants with targeted genes in the filamentous fungus *Leptosphaeria maculans*. Curr. Genet. 45: 245-249.

Giancola S., Marhadour S., Desloire S., Clouet V., Falentin-Guyomarc'h H., Laloui W., Falentin C., Pelletier G., Renard M., Bendahmane A., Delourme R., Budar F. (2003) Characterization of a radish introgression carrying the Ogura fertility restorer gene *Rfo* in rapeseed, using the Arabidopsis genome sequence and radish genetic mapping. Theor Appl Genet. 107: 1442-1451.

Girke T., Todd J., Ruuska S., White J., Benning C., Ohlrogge J. (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol. 124: 1570-1581.

Government of Canada. Canadian Food Inspection Agency. List of Varieties Which are Registered in Canada. Released July 1st, 2004. http://www.inspection.gc.ca/english/plaveg/variet/lovric.pdf

Hammond K.E., Lewis B.G. (1987a) The establishment of systemic infection in leaves of oilseed rape by *Leptosphaeria maculans*. Plant Pathology 36: 135-147.

Hammond K.E., Lewis B.G. (1987b) Variation in steminfections caused by aggressive and non-aggressive isolates of *Leptosphaeria maculans* on *Brassica napus* var. *oleifera*. Plant Pathology 36: 53-65.

Hammond-Kosack K.E., Jones J.D.G. (1996) Resistance gene dependent plant-defense responses. Plant Cell 8:1773-1791.

Howlett B.J., Idnurm A., Pedras M.S. (2001) *Leptosphaeria maculans*, the causal agent of blackleg disease of *Brassicas*. Fungal Genet Biol. 33: 1-14.

Idnurm A., Howlett B.J. (2002) Isocitrate lyase is essential for pathogenicity of the fungus *Leptosphaeria maculans* on canola (Brassica napus). Eukaryotic Cell 1: 719-724.

Idnurm A., Howlett B.J. (2003) Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. Fungal Genet Biol. 39: 31-7.

Jia Y., McAdams S.A., Bryan G.T., Hershey H.P, Valent B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J* 19: 4004–4014.

Johal G.S., Briggs S.P. (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. Science. 258: 985-7.

Keri M. (1999) Genetic studies of host–pathogen interaction between *Brassica napus* and *Leptosphaeria maculans*. Ph.D. thesis. University of Manitoba Winnipeg, Manitoba, Canada.

Keri M., van den Berg C.J.G., McVetty P.B.E., Rimmer S.R. (1997) Inheritance of resistance to *Leptosphaeria maculans* in *Brassica juncea*. Phytopathology 87: 594-598.

Koch E., Song K., Osborn T.C., Williams P.H. (1991) Relationship between pathogenicity and phylogeny based on restriction fragment length polymorphism in *Leptosphaeria maculans*. Mol. Plant-Microbe Interact. 4: 341-349.

Krysan P.J., Young J.C., Sussman M.R. (1999) T-DNA as an insertional mutagen in Arabidopsis. Plant Cell 11: 2283-2290.

Kunkel B.M., Brooks D.M. (2002) Crosstalk between signaling pathways in pathogen defense. Current Opinion in Plant Biology 5: 325-331.

Kuswinanti T., Sock J., Hoppe H. H. (1995) Variation in virulence of aggressive isolates of *Leptosphaeria maculans* based on cotyledon reactions on an extended differential set. Proc. Int. Rapeseed Congr., 9th. Cambridge, U.K. Pages 1248-1250.

Lagercrantz U., Lydiate D.J. (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. Genome 38: 255-264.

Leclair S., Ansan-Melayah D., Rouxel T., Balesdent M. H. (1996) Meiotic behaviour of the minichromosome in the phytopathogenic ascomycete *Leptosphaeria maculans*. Curr. Genet. 30: 541-548.

Li X., Zhang Y. (2002) Reverse genetics by fast neutron mutagenesis in higher plants. Funct. Integr. Genomics 2: 254-258

Lim L., Howlett B.J. (1994) Linear plasmids, pLm9 and pLm10, can be isolated from the phytopathogenic ascomycete *Leptosphaeria maculans* by pulsed-field gel electrophoresis. Curr. Genet. 26: 276-280.

Mayerhofer R., Bansal V.K., Thiagarajah M.R., Stringam G.R., Good A.G. (1997) Molecular mapping of resistance to *Leptosphaeria maculans* in Austrailian cultivars of *Brassica napus*. Genome, 40: 294-301.

Mackey D., Belkhadir Y, Alonso J.M., Ecker J.R., Dangl J.L. (2003) Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. Cell 112: 379-389.

Melchers L.S., Stuiver M.H. (2000) Novel genes for disease-resistance breeding. Curr Opin Plant Biol. 3: 147-152.

Morales V.M., Seguin-Swartz G., Taylor J.L. (1993) Chromosome size polymorphism in *Leptosphaeria maculans.* Phytopathology 83: 503-509.

Narusaka Y, Narusaka M, Seki M, Ishida J, Nakashima M, Kamiya A, Enju A, Sakurai T, Satoh M, Kobayashi M, Tosa Y, Park P, Shinozaki K. (2003) The cDNA microarray analysis using an Arabidopsis pad3 mutant reveals the expression profiles and classification of genes induced by *Alternaria brassicicola* attack. Plant Cell Physiol. 44: 377-387.

Nimchuk Z., Rohmer L., Chang J.H. Dangl J. (2001) Knowing the dancer from the dance: *R*-gene products and their interactions with other proteins from host and pathogen. Current Opinion in Plant Biology 4:288-294.

Pang, E.C.K. Halloran, G.M. (1996) The genetics of adult-plant blackleg (*Leptosphaeria maculans*) resistance from *Brassica juncea* in *B. napus*. Theor Appl Genet 92: 382-387.

Parkin I.A., Sharpe A.G., Keith D.J., Lydiate D.J. (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome. 38: 1122-1131.

Parkin I.A., Lydiate D.J., Trick M. (2002) Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. Genome. 45: 356-366.

Pedras M.S.C., Taylor J.L., Morales V.M. (1996) The blackleg fungus of rapeseed: How many species? Acta-Hortic. 441-446.

Perez-Garcia A., Snoeijers S.S., Joosten M.H., Goosen T., De Wit P.J. (2001) Expression of the Avirulence gene *Avr9* of the fungal tomato pathogen *Cladosporium fulvum* is regulated by the global nitrogen response factor NRF1. Mol. Plant Microbe Interact. 14: 316-325.

Perfect S.E., O'Connell R.J., Green E.F., Doering-Saad C., Green J.R. (1998) Expression cloning of a fungal proline-rich glycoprotein specific to the biotrophic interface formed in the Colletotrichum-bean interaction. Plant J. 15: 273-279.

Petrie G.A. (1978) Occurrence of a highly virulent strain of blackleg (*Leptosphaeria maculans*) on rape in Saskatchewan (1975-77). Can. Plant Dis. Surv. 65: 35-41.

Plieske J., Struss D. Robbelen G. (1998) Inheritance of resistance derived from the B-genome of *Brassica* against *Phoma lingam* in rapeseed and the development of molecular markers. Theor Appl Genet 97: 929-936.

Plieske J. Struss D. (2001) STS markers linked to *Phoma* resistance genes of the *Brassica* B-genome revealed sequence homology between *Brassica nigra* and *Brassica napus*. Theor Appl Genet. 102: 483-488.

Plummer K.M., and Howlett B.J. (1993) Major chromosomal length polymorphisms are evident after meiosis in the phytopathogenic fungus *Leptosphaeria maculans*. Curr Genet. 24: 107-113.

Plummer KM, Howlett BJ. (1995) Inheritance of chromosomal length polymorphisms in the ascomycete *Leptosphaeria maculans*. Mol Gen Genet. 247: 416-22.

Pongam P., Osborn T. C., Williams P. H. (1998) Genetic analysis and identification of amplified fragment length polymorphism markers linked to the *alm1* avirulence gene of *Leptosphaeria maculans*. Phytopathology 88: 1068-1072.

Purwantara A., Salisbury P.A., Burton W.A., Howlett B.J. (1999) Reaction of *Brassica juncea* (Indian mustard0 lines to Austrailian isolates of *Leptosphaeria maculans* under glasshouse and field conditions. Eur J Plant Path 104(9): 895-902.

Rhee S.Y., Beavis W., Berardini T.Z., Chen G., Dixon D., Doyle A., Garcia-Hernandez M., Huala E., Lander G., Montoya M., Miller N., Mueller L.A., Mundodi S., Reiser L., Tacklind J., Weems D.C., Wu Y., Xu I., Yoo D., Yoon J., Zhang P. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Research 31: 224-228.

Rimmer S.R., van den Berg C.G.J. (1992) Resistance of oilseed *Brassica* spp. To blackleg caused by *Leptosphaeria maculans*. Can J Plant Path 14: 56-66.

Rouxel T., Willner E,. Coudard L., Balesdent M.H. (2003) Screening and identification of resistance to *Leptosphaeria maculans* (stem canker) in *Brassica napus* accessions. Euphytica 133: 219-231.

Schenk P.M., Kazan K.,Wilson I., Anderson J.P., Richmond T., Somerville S.C., Manners J.M. (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. PNAS 97: 11655-11660.

Sessions A., Burke E., Presting G., Aux G., McElver J., Patton D., Dietrich B., Ho P, Bacwaden J., Ko C., Clarke J.D., Cotton D., Bullis D., Snell J., Miguel T., Hutchison D., Kimmerly B., Mitzel T., Katagiri F., Glazebrook J., Law M., Goff S.A. (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell. 14: 2985-2994.

Sexton A.C., Howlett B.J. (2000) Characterisation of a cyanide hydratase gene in the phytopathogenic fungus *Leptosphaeria maculans*. Mol Gen Genet. 263: 463-470.

Sexton A.C., Paulsen M., Woestemeyer J., Howlett B.J. (2000) Cloning, characterization and chromosomal location of three genes encoding host-cell-wall-degrading enzymes in *Leptosphaeria maculans*, a fungal pathogen of *Brassica* spp. Gene. 248: 89-97.

Shoemaker R.A., Brun H. (2001) The teleomorph of the weakly aggressive segregate of *Leptosphaeria maculans*. Can. J. Bot. 79: 412-419.

Sillito D., Parkin I.A., Mayerhofer R., Lydiate D.J., Good A.G. (2000) *Arabidopsis thaliana*: a source of candidate disease-resistance genes for *Brassica napus*. Genome. 43: 452-460.

Solomon P.S., Nielsen P.S., Clark A.J., Oliver R.P. (2000) Methionine synthase, a gene required for methionine synthesis, is expressed in planta by *Cladosporium fulvum*. Mol. Plant Pathol. 1: 315-323.

Solomon P.S., Oliver R.P. (2001) The nitrogen content of the tomato leaf apoplast increases during infection by *Cladosporium fulvum*. Planta 213: 241-249.

Solomon P.S., Oliver R.P. (2002) Evidence that gamma-aminobutyric acid is a major nitrogen source during *Cladosporium fulvum* infection of tomato. Planta. 214: 414-420.

Somda I., Harkous S., Brun H. (1997) Bipolar heterothallism in B-group isolates of *Leptosphaeria maculans*. Plant Pathol. 46: 890-896.

Taylor J.L., Borgmann I., Seguin-Swartz G. (1991) Electrophoretic karyotyping of *Leptosphaeria maculans* differentiates highly virulent from weakly virulent isolates. Curr. Genet. 19: 273-278.

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 408: 796-815.

Warwick S.I., Black L.D., Aguinagalde I. (1992) Molecular systematics of *Brassica* and allied genera (Subtribe Brassicinae, Brassiceae) -- chloroplast DNA variation in the genus *Diplotaxis*. Theor Appl Genet. 83: 839-850.

Williams R.S., Davis M.A., Howlett B.J. (1994) Nitrate reductase of the ascomycetous fungus, *Leptosphaeria maculans:* gene sequence and chromosomal location. Mol. Gen. Genet. 244: 1-8.

Woods, D.L., Capcara, J.J., Downey, R.K. (1991) The potential of mustard (*Brassica juncea* (L.) Coss.) as an edible oil crop on the Canadian prairies. Can. J. Plant Sci. 71: 195-198.

Woods, D.L., Capcara, J.J., Downey, R.K. (1991) The potential of mustard (*Brassica juncea* (L.) Coss.) as an edible oil crop on the Canadian prairies. Can. J. Plant Sci. 71: 195-198.

Wretblad S., Bohman S., Dixelius C. (2003) Overexpression of a *Brassica nigra* cDNA gives enhanced resistance to *Leptosphaeria maculans* in *B. napus*. Mol. Plant-Microbe Interact. 16: 477-484.

Yang Y.W., Lai K.N., Tai P.Y., Li W.H. (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J. Mol. Evol. 48: 597-604.

U. N. (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. Jpn. J. Bot. 7: 389-452.

2. Genetic mapping of genes for resistance to *Leptosphaeria maculans* in *Brassica juncea*

2.1. Introduction.

Brassica species that contain the B genome, i.e. *Brassica nigra* (BB), *Brassica juncea* (AABB), and *Brassica carinata* (BBCC), are known to have high levels of blackleg resistance (reviewed in Rimmer and van den Berg 1992). Many attempts to exploit these resistance sources for use in the more extensively cultivated and more susceptible crop species *Brassica napus* have been mounted over the last two and a half decades. Transfer of resistance from *B. juncea* to *B. napus* through the creation of interspecific hybrids was first described by Roy (1978). Early cytological studies of these hybrid derived resistant lines indicated that they possessed the correct number of chromosomes for *B. napus* (2n=38), although true breeding resistance could not be obtained (Roy 1984). Later examinations of lines developed from this material found that the majority were actually anueploid (Rimmer and van den Berg 1992). Material from this and other interspecific hybrids has been used in attempts to characterize the inheritance of *B. juncea* resistance to *Leptosphaeria maculans*. In crosses between lines derived from Roy's hybrids and a susceptible *B. napus* variety, one dominant and two recessive genes each capable of conferring resistance were identified (Pang and Halloran 1996). All of the F3 progeny examined from these crosses had the normal diploid number of chromosomes for *B. napus*. The *B. napus* (AACC) parent used in the initial production of these hybrid plants (but not the study population) is also resistant to the blackleg strains used, calling into question the source of these genes for resistance. Another study using material from Roy's hybrids reported finding one dominant gene for resistance to blackleg, which was shown by random amplified polymorphic DNA (RAPD) marker analysis to originate from the B genome (Chevre et al. 1997). Establishment of the line used in this study required cytological selection for appropriate chromosome number and regular meiotic behaviour.

Hybrid plants have also been generated from crosses between *B. napus* and each of the three B genome containing Brassica species; *B. juncea, B. nigra* and *B. carinata* (Plieske et al. 1998). In each case offspring from these hybrid plants contained a single gene conferring resistance to blackleg. Plieske and Struss (2001) reported that markers linked to these resistance traits were mapped to linkage group G3 of the *B. nigra* map of Lagercrantz and Lydiate (1995), which is equivalent to linkage group J13 of the *B. juncea* map (Axelsson et al. 2000).

Asymmetric somatic hybridizations between *B. nigra* and *B. napus* as well as *B. juncea* and *B. napus* have also been produced (Dixelius 1999; Dixelius and Wahlberg 1999). Three different resistance loci were identified in these populations. Each locus co-segregated with the B genome allele of a different RFLP marker. These markers were located on linkage groups G2, G5 and G8.

Comparisons between these various studies are complicated by the use of different B genome donor material, different recurrent parents in the various crossing strategies, different pathogen isolates or isolate mixtures, lack, or sparseness of genetic marker data and the use of different marker sets. All previous examinations of resistance to blackleg in *B. juncea* have suffered from the disadvantage of using hybrid genomes to move resistance traits from *B. juncea* cultivars to susceptible *B. napus* cultivars for study. Large genomic introgressions tend not to recombine readily with their host genomes, resulting in large amounts of linkage drag associated with a trait of interest, as is the case with the *ogura* fertility restorer gene introgressed from *Raphanus sativa* into *B. napus* (Delourme et al. 1998). This has negative implications for breeding, as undesirable foreign genome elements are difficult to dissociate from the trait of interest. This linkage drag also renders map based cloning very impractical as any markers on an introgressed genome fragment will appear very tightly linked genetically to the trait of interest despite being very far apart physically.

Though largely resistant to *L. maculans,* susceptible varieties of *B. juncea* have been described. Two highly virulent PG4 Australian isolates of *L. maculans* have been demonstrated to cause disease on all *B. juncea* lines tested (27 in the greenhouse, 92 in the field) (Purwantara et al. 1999). Also, three *B. juncea* lines of 296 tested were identified that gave susceptible reactions when inoculated with PG2 isolates of *L. maculans* collected from Manitoba (Keri 1991). The use of these susceptible accessions to study *B. juncea* resistance in natural populations avoids difficulties with meiotic instability and makes possible genetic mapping at normal resolution.

This thesis reports the crossing of the susceptible accession of *B. juncea* described by Keri (1991) with AC Vulcan, a blackleg resistant *B. juncea* mustard variety, to generate F2 and backcross 1 (B1) mapping populations segregating for resistance to *L. maculans* infection. Segregation for resistance in the F2 population and selected F3 families identified the presence of two genes conferring resistance, one dominant and one recessive in nature, as predicted by Keri (1997). A detailed linkage map was constructed from the F2 population using both RFLP and microsattelite markers. The dominant gene for resistance to *L. maculans*, *LMJR1*, was mapped on the linkage group J13 of the *B. juncea* B genome.

2.2. Methods and materials

2.2.1. Parental lines and population development.

AC Vulcan is a *B. juncea* mustard cultivar developed at Agriculture and Agri-Food Canada (Gerhard Rakow; AAFC, Saskatoon SK, Canada). UM3132 is an inbred line of *B. juncea* described by Keri (1991). Individual AC Vulcan plants were pollinated by UM3132. A single F1 plant from this cross was selfed to create an F2 population segregating for resistance. This same F1 individual was also used to pollinate the susceptible parent UM3132 to generate a backcross 1 (B1) population (Figure 2-1).

2.2.2. Preparation of *L. maculans* inoculum

L. maculans isolate PL86-12 collected in Manitoba and made available from the collection at the AAFC Saskatoon Research Centre was prepared by culturing surface sterilized, infected Westar cotyledons on V-8 agar (20% Campbells V-8 juice, 0.075% calcium carbonate and 100μg/l streptomycin sulfate) plates. Plates were incubated at room temperature under continuous fluorescent light. After 2 weeks, spores were collected in sterile water. Spore suspensions were filtered through sterilized whatman filter paper and the concentration was adjusted to 1x 10^7 spores/ml in sterile distilled water. Spore suspensions were stored at -20°C.

2.2.3. Plant cultivation and disease ratings.

Seedlings were germinated in soil in 36 well flats at 18-24°C with a photoperiod of 14-18 hours. At the emergence of the first true leaves, cotyledons were inoculated by puncturing each lobe with a 21 gauge needle (4 punctures/plant, two/cotyledon). A 10 ml droplet of spore suspension (1x 10^7 spores/ml) was deposited over each puncture and allowed to dry. Plants were kept in normal greenhouse conditions and rated at 14 days post inoculation. Disease severity was rated according to the 9 point rating scale devised by Delwiche (1980), where the numerical value of the rating increases with the severity of the infection 0= no symptoms to 9= very susceptible. After rating, plants were repotted into 4 inch square plots and grown for tissue and seed collection.

2.2.4. Marker analysis using restriction fragment length polymorphism (RFLP) markers

DNA extraction and southern hybridization were performed as described in Sharpe et al. (1995). 10 µg of genomic DNA from each sample was digested with one of *Eco* RI, *Eco* RV or *Bam* HI. DNA fragments were size separated by gel electrophoresis and gels were blotted using a standard alkali transfer protocol with Hybond N nylon membranes (Amersham Biosciences). Southern blots of genomic DNA were probed with *Pst I* fragments of genomic DNA (pW, pN, pO and pR probes) used previously to detect RFLPs in *Brassica* species (Sharpe et al. 1995; Axellson et al. 2001).

2.2.5. Microsatellite markers

Highly informative *B. juncea* and *B. napus* microsatellite markers with defined loci in the

Brassica A and B genomes have been developed by D.J. Lydiate and A.G. Sharpe at AAFC Saskatoon Research Centre, Saskatoon SK, Canada. A subset of these markers was used to characterize the *B. juncea* populations. Amplification of microsatellite repeats was performed using an ABI 877 thermocycler/liquid handling robot (Applied Biosystems) or the MWG Primus HT multiblock thermocycler (MWG Biotech). Microsatellite regions were amplified using specific paired primers (0.33 μM), one of which was labeled with a fluorescent dye (Rox, Fam, or Tet). Amplification reactions were carried out with 40ng of genomic DNA, 0.2 mM dNTPs, 2.5mM MgCl2, and 0.5U Amplitaq Gold (Applied Biosystems) under the following conditions: 95 °C for 10 min.; seven cycles of 94 ° C for 15s; 50 °C for 15s; 72 °C for 30s; 27 cycles of 89 °C for 15s; 50 ° for 15s; 72 ° for 30s and a final extension step of 72 ° for 10 min. The majority of reactions were multiplexed. Products from six reactions (two reactions per fluorescent dye) were pooled and separated through polyacrylimide gels on an ABI 377 automated DNA sequencer (Applied Biosystems).

2.2.6. Genetic mapping

Genetic linkage analysis was performed using MAPMAKER 3.0 (Lander et al. 1987). A minimum LOD score of 4.0 and a maximum distance of 50.0 cM were used for the initial linkage association. A LOD score of 3.0 was used to bridge the four largest gaps. Ordering markers within linkage groups was performed using the order command of MAPMAKER. Scorings were checked in instances where double crossovers flanked single loci and the map order was recalculated following all corrections. Recombination frequencies were converted to Kosambi Centimorgan map distances (Kosambi 1944). The significance of segregation distortion at each marker locus was tested by comparing

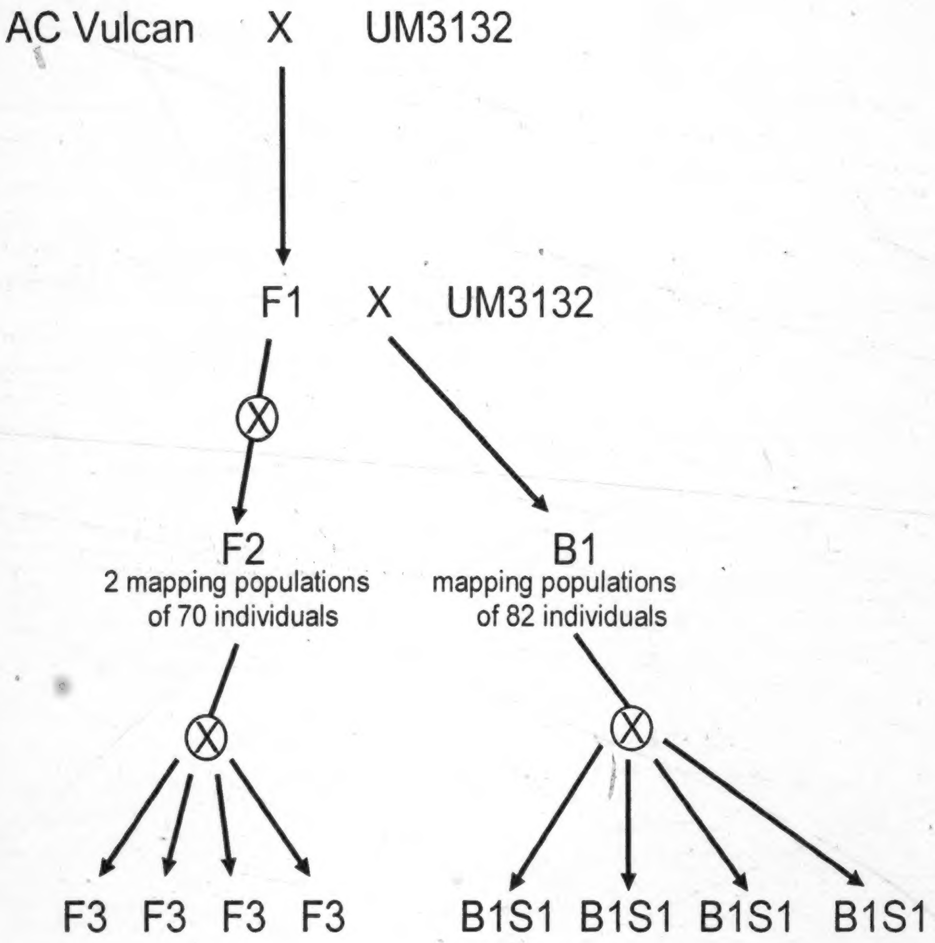

AC Vulcan X UM3132
F1 X UM3132
F2
2 mapping populations of 70 individuals
B1
mapping populations of 82 individuals
F3 F3 F3 F3
B1S1 B1S1 B1S1 B1S1

Figure 2-1. Crossing strategy and population development. A single plant of *Brassica juncea* cultivar AC Vulcan was pollinated by a single plant of *B. juncea* accession UM3132. A single F1 plant was selfed to produce an F2 population segregating for resistance to blackleg. This same F1 individual was used to pollinate UM3132 to produce a first backcross (B1) population. Select individuals of both the F2 and B1 populations were also selfed to yield F3 and B1S1 families.

the observed number of individuals that were homozygous for AC Vulcan alleles, heterozygous or homozygous for UM3132 alleles against the expected numbers based on a theoretical 1:2:1 segregation ratio and calculating a chi-squared statistic and p-value.

2.3. Results

2.3.1 Production and phenotypic analysis of mapping populations

Each of the parents used in this study was tested for resistance phenotype before being used to generate mapping populations. An early candidate for the resistant parent in the crosses, J-O-7-5, a line derived from J-O-7 was discarded as it was not uniformly resistant. Of 19 plants tested, 7 rated susceptible, the mean disease rating for this sample was 4.37. Progeny from these plants also segregated for resistance. The cause of the non-uniformity of resistance responses in the J-O-7-5 accession used was not investigated. It is possible to speculate on reasons for the appearance of susceptible types. These possibilities include rare genetic events such as a translocation near the resistance gene, although causes such as seed or pollen contamination of the J-O-7-5 stock by UM3132 cannot be ruled out. An alternative resistant parent, AC Vulcan, was selected. AC Vulcan had a mean disease rating of 1.78, although one individual scored a 7 which would class as susceptible. UM3132 had a mean disease rating of 8.79 (Figure 2-2A). The AC Vulcan X UM3132 F1 was uniformly resistant ranging from 1-3 with one individual scoring a 5, the mean score was 1.53 (data not shown). One arbitrarily chosen individual from this group of F1s was used to pollinate UM3132 producing backcross offspring and selfed to produce a segregating F2 population (Figure 2-2B). A total of 178 plants were grown and tested in the F2 population. Initially, a ratio of 120 resistant (0-3 rating) to 16 intermediate (4-6) to 42 susceptible (7-9) plants was observed. Families of F3 progeny from selfed F2 individuals representing each of the three classes were grown and rated (Table 2-1). All 10 F3 families from resistant F2 confirmed the initial disease rating. Of the 11 F3 families tested from the intermediate class, 9 indicated the F2 parent plant contained a resistance trait, the other two indicated that the F2 parent was, in fact susceptible. A surprisingly high number of F3 families from susceptible F2 parents were completely or predominantly resistant (10 of 25), revealing that the initial F2 rating was likely mistaken. Progeny were tested from all F2 susceptible individuals that yielded enough seed to do so. Of the 15 confirmed susceptible F3 families, 9 were segregating for resistance consistent with the 1 resistant to 3 susceptible plants ratio expected in the two gene model predicted by Keri (1997), while the other 6 bred true for susceptibility. F2 phenotypic scorings were adjusted to reflect any changes necessitated by segregation for resistance in F3 progeny. The revised segregation ratio for resistance to susceptibility is 144:34. This ratio could fit a single gene 3:1 model (χ^2= 3.033, p-value= 0.082, df= 1) but also fits a 13:3 two gene model where one gene is dominant and the other recessive (χ^2= 0.0144, p-value= 0.90, df= 1), although the number of susceptible individuals in this classification may be slightly inflated as it can be expected that some of the untested susceptible F2 plants are in fact resistant.

2.3.2. Developing a map of the *B. juncea* genome

A combination of microsatellite and RFLP markers was used to generate a genetic map of *B. juncea* (Figure 2-3). As genomic DNA from some plants was limiting, the

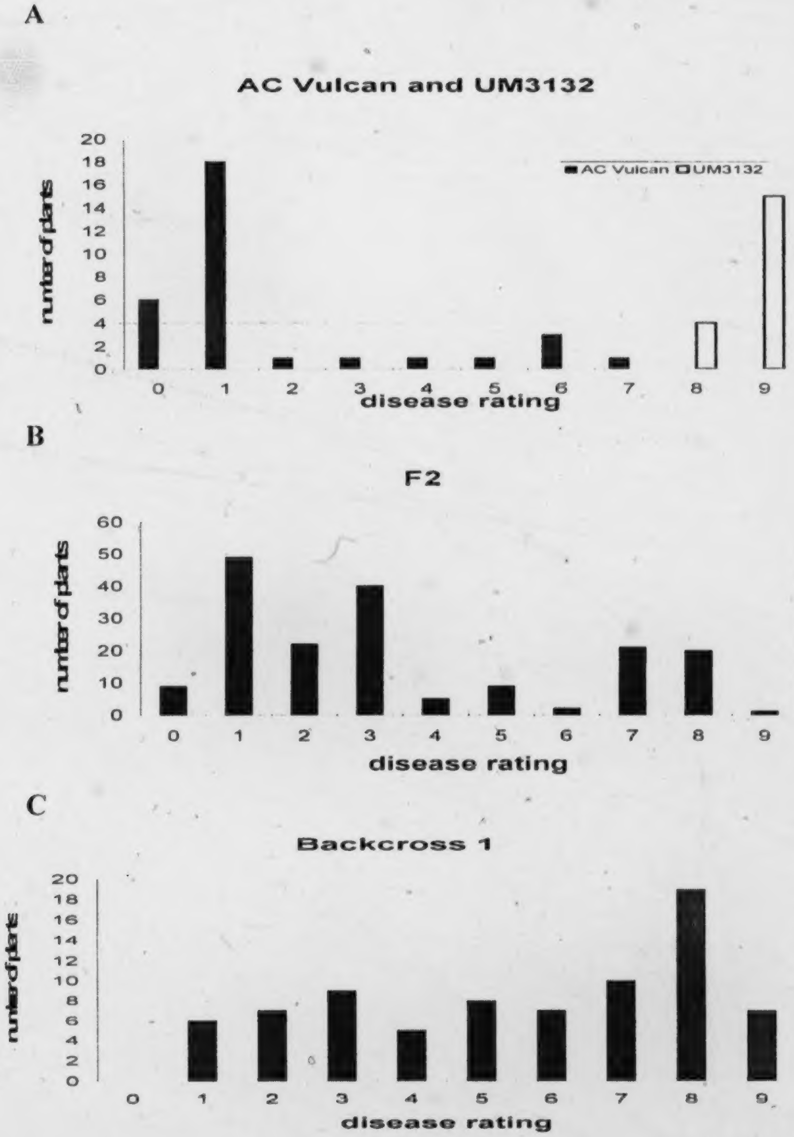


Figure 2-2. Disease ratings for parental genotypes and plants of the segregating F2 and B1 populations. The graphs represent the frequency distributions of the degree of susceptibility to infection by *L. maculans* strain PL 86-12 (using the numerical ratings described by Delwiche (1980)) exhibited by plant populations tested: **A**, plants of AC Vulcan (the resistant parent) and UM3132 (the susceptible parent); **B** plants of the F2 generation segregating for resistance to blackleg; **C** plants of the B1 generation segregating for resistance.

Table 2-1. Segregation for resistance in F3 progeny derived from F2 individuals scoring intermediate or susceptible.

initial rating	F2	F3 segregation		best fit	chi-square	p-value	Final rating
		R	S				
I	23	20	0	1:0			R
I	33	13	6	3:1	0.44	0.51	R
I	56	21	0	1:0			R
I	2	15	0	1:0			R
I	7	19	0	1:0			R
I	19	17	2	3:1	2.12	0.15	R
I	34	12	5	3:1	0.18	0.67	R
I	43	11	4	3:1	0.02	0.88	R
I	155	11	5	3:1	0.33	0.56	R
I	42	8	12	1:3	2.40	0.12	S
I	44	6	10	1:3	1.33	0.25	S
S	27	7	5	3:1	1.78	0.18	R
S	97	18	2	3:1	2.40	0.12	R
S	55	11	2	3:1	0.64	0.42	R
S	1	10	5	3:1	0.56	0.46	R
S	6	10	0	1:0			R
S	26	12	7	3:1	1.42	0.23	R
S	40	14	6	3:1	0.27	0.61	R
S	108	10	8	3:1	3.63	0.06	R
S	145	15	4	3:1	0.16	0.69	R
S	169	20	0	1:0			R
S	28	3	5	1:3	0.67	0.41	S
S	62	2	16	1:3	1.85	0.17	S
S	102	4	10	1:3	0.10	0.76	S
S	107	1	16	1:3	3.31	0.07	S
S	109	0	11	0:1			S
S	114	0	5	0:1			S
S	131	0	12	0:1			S
S	139	2	14	1:3	1.33	0.25	S
S	36	9	12	1:3	3.57	0.06	S
S	101	6	10	1:3	1.33	0.25	S
S	130	1	17	1:3	3.63	0.06	S
S	140	3	11	1:3	0.10	0.76	S
S	144	6	14	1:3	0.27	0.61	S
S	161	8	12	1:3	2.40	0.12	S
S	143	0	20	0:1			S
R	24	12	8	3:1	2.40	0.12	R
R	77	20	0	1:0			R
R	18	17	3	3:1	1.07	0.30	R
R	80	19	1	3:1	4.27	0.04	R
R	20	16	2	3:1	1.85	0.17	R
R	65	15	5	3:1	0.00	1.00	R
R	125	10	8	3:1	3.63	0.06	R
R	168	17	3	3:1	1.07	0.30	R
R	58	20	0	1:0			R
R	61	14	6	3:1	0.27	0.61	R

original population of 140 individuals was divided into two subpopulations. Each subpopulation was screened with a different set of RFLP markers, producing two maps. Both subpopulations were screened with all microsattelite markers used. Microsatellite marker loci were then used as bridges to integrate the two RFLP maps. The map for subpopulation 1 consists of 257 loci across 18 linkage groups, 1 unassigned triplet, 1 unassigned doublet and 14 independent unlinked loci. Total map coverage is 1577.3 Kosambi cM. The subpopulation 2 map consists of 113 loci across 16 linkage groups, 3 unassigned triplets, 4 unassigned doublets and 26 independent unlinked loci. Total map coverage is 721.4 Kosambi cM. When combined, the map for subpopulation 2 overlaps regions present in the subpopulation 1 map and does not extend the map coverage significantly. The map described here corresponds almost exactly to a previously described genetic map of *B. juncea* (Axelsson et al. 2000), save for six pairs of markers that are inverted between the two maps. A-genome and B-genome linkage groups could be identified on the basis of comparison to this map. J1 to J10 comprise the A genome present in *B. juncea* while J11 to J18 make up the B genome. The four largest intervals, which could not be bridged at LOD 4.0, have previously been demonstrated to be linked in the existing *B. juncea* map (Axelsson et al. 2000).

Marker scorings for each locus were tested for segregation distortion using a chi-squared goodness-of-fit test against the expected 1:2:1 allele ratio. On linkage groups J2-J18, 17 loci were significantly distorted ($p< 0.05$), a further 6 loci were significantly distorted ($p <0.01$). Loci with distorted segregation are indicated on the map (Figure 2-3).

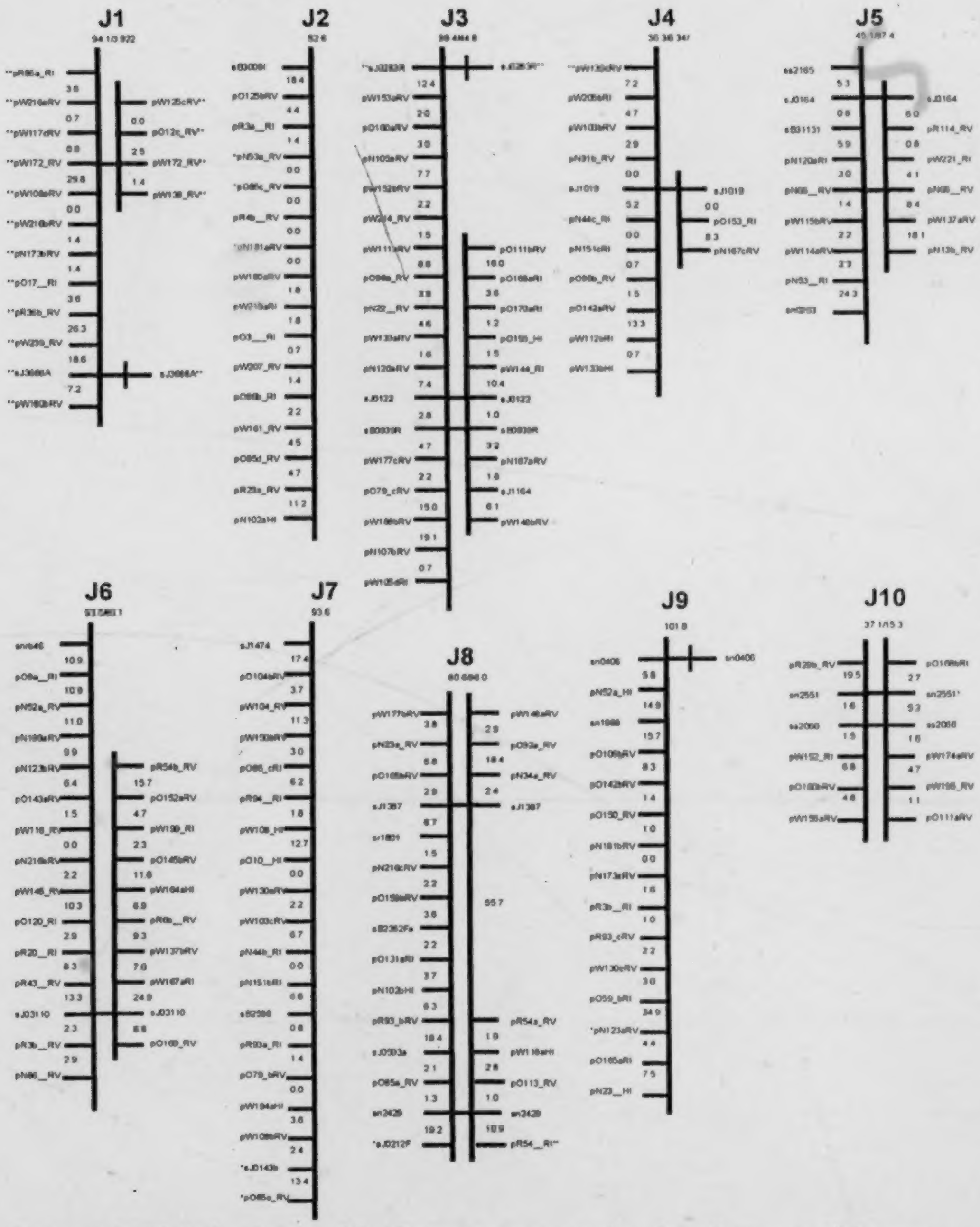

J1
94 1/3 922
**pR85a_RI
3.8
**pW216aRV
0.7
**pW117cRV
0.8
**pW172_RV
29.8
**pW108aRV
0.0
**pW216bRV
1.4
**pN173bRV
1.4
**pO17__RI
3.6
**pR36b_RV
26.3
**pW239_RV
18.6
**sJ3888A
7.2
**pW160bRV
pW125cRV*
0.0
pO12c_RV*
2.5
pW172_RV*
1.4
pW138_RV*
sJ3888A**
J2
52.6
sB3008I
18.4
pO125bRV
4.4
pR3a__RI
1.4
*pN53a_RV
0.0
*pO88c_RV
0.0
pR4b__RV
0.0
*pN181aRV
0.0
pW180aRV
1.8
pW218aRI
1.8
pO3___RI
0.7
pW207_RV
1.4
pO80b_RI
2.2
pW161_RV
4.5
pO85d_RV
4.7
pR29a_RV
11.2
pN102aHI
J3
99.4/44.8
**sJ0283R
12.4
pW153aRV
2.0
pO160aRV
3.0
pN105aRV
7.7
pW152bRV
2.2
pW214_RV
1.5
pW111aRV
8.6
pO98a_RV
3.8
pN22__RV
4.6
pW133aRV
1.6
pN120aRV
7.4
sJ0122
2.8
sB0939R
4.7
pW177cRV
2.2
pO79_cRV
15.0
pW186bRV
19.1
pN107bRV
0.7
pW105dRI
sJ0283R**
pO111bRV
16.0
pO168aRI
3.6
pO170aRI
1.2
pO155_HI
1.5
pW144_RI
10.4
sJ0122
1.0
sB0939R
3.2
pN167aRV
1.8
sJ1164
6.1
pW140bRV
J4
36.38.34/
**pW139dRV
7.2
pW205aRI
4.7
pW103bRV
2.9
pN31b_RV
0.0
sJ1019
5.2
pN44c_RI
0.0
pN151cRI
0.7
pO98b_RV
1.5
pO142aRV
13.3
pW112bRI
0.7
pW133bHI
sJ1019
0.0
pO153_RI
8.3
pN167cRV
J5
45.1/87.4
ss2165
5.3
sJ0164
0.8
sB31131
5.9
pN120aRI
3.0
pN88__RV
1.4
pW115bRV
2.2
pW114aRV
2.2
pN53__RI
24.3
sn0203
sJ0164
6.0
pR114_RV
0.8
pW221_RI
4.1
pN88__RV
8.4
pW137aRV
18.1
pN13b_RV
J6
93.0/83.1
snrb46
10.9
pO8a__RI
10.8
pN52a_RV
11.0
pN180aRV
9.9
pN123bRV
6.4
pO143aRV
1.5
pW116_RV
0.0
pN216bRV
2.2
pW145_RV
10.3
pO120_RI
2.9
pR20__RI
8.3
pR43__RV
13.3
sJ03110
2.3
pR3b__RV
2.9
pN86__RV
pR54b_RV
15.7
pO152aRV
4.7
pW189_RI
2.3
pO145bRV
11.8
pW164aHI
6.9
pR6b__RV
9.3
pW137bRV
7.8
pW167aRI
24.9
sJ03110
6.8
pO169_RV
J7
93.6
sJ1474
17.4
pO104bRV
3.7
pW104_RV
11.3
pW150aRV
3.0
pO86_cRI
6.2
pR94__RI
1.8
pW108_HI
12.7
pO10__HI
0.0
pW130aRV
2.2
pW103cRV
6.7
pN44b_RI
0.0
pN151bRI
6.6
sB2598
0.8
pR93a_RI
1.4
pO79_bRV
0.0
pW194aHI
3.6
pW108bRV
2.4
*sJ0143b
13.4
*pO85e_RV
J8
80.6/96.0
pW177bRV
3.8
pN23a_RV
6.8
pO105bRV
2.9
sJ1387
8.7
sr1801
1.5
pN216cRV
2.2
pO159aRV
3.6
sB2352Fa
2.2
pO131aRI
3.7
pN102bHI
6.3
pR93_bRV
18.4
sJ0503a
2.1
pO85a_RV
1.3
sn2429
19.2
*sJ0212F
pW146aRV
2.8
pO92a_RV
18.4
pN34a_RV
2.4
sJ1387
95.7
pR54a_RV
1.9
pW118aHI
2.8
pO113_RV
1.0
sn2429
18.9
pR54__RI**
J9
101.8
sn0406
5.8
pN52a_HI
14.9
sn1988
15.7
pO109bRV
8.3
pO142bRV
1.4
pO150_RV
1.0
pN181bRV
0.0
pN173aRV
1.6
pR3b__RI
1.0
pR93_cRV
2.2
pW130cRV
3.0
pO59_bRI
34.9
*pN123aRV
4.4
pO165aRI
7.5
pN23__HI
sn0406
J10
37.1/15.3
pR29b_RV
19.5
sn2551
1.6
ss2050
1.5
pW152_RI
6.8
pO180bRV
4.8
pW155aRV
pO168bRI
2.7
sn2551*
5.2
ss2086
1.6
pW174aRV
4.7
pW198_RV
1.1
pO111aRV

Figure 2-3. F2 derived map of the *B. juncea* genome showing alignment of the two subpopulation maps. Linkage groups J1 through J10 belong to the A genome, groups J11 through J18 belong to the B genome. The length of each linkage group, in Kosambi centimorgans is listed beneath the linkage group title. The left hand vertical bars represent the subpopulation 1 map; right hand vertical bars represent the subpopulation 2 map. Horizontal lines between the bars indicate loci common to both maps.

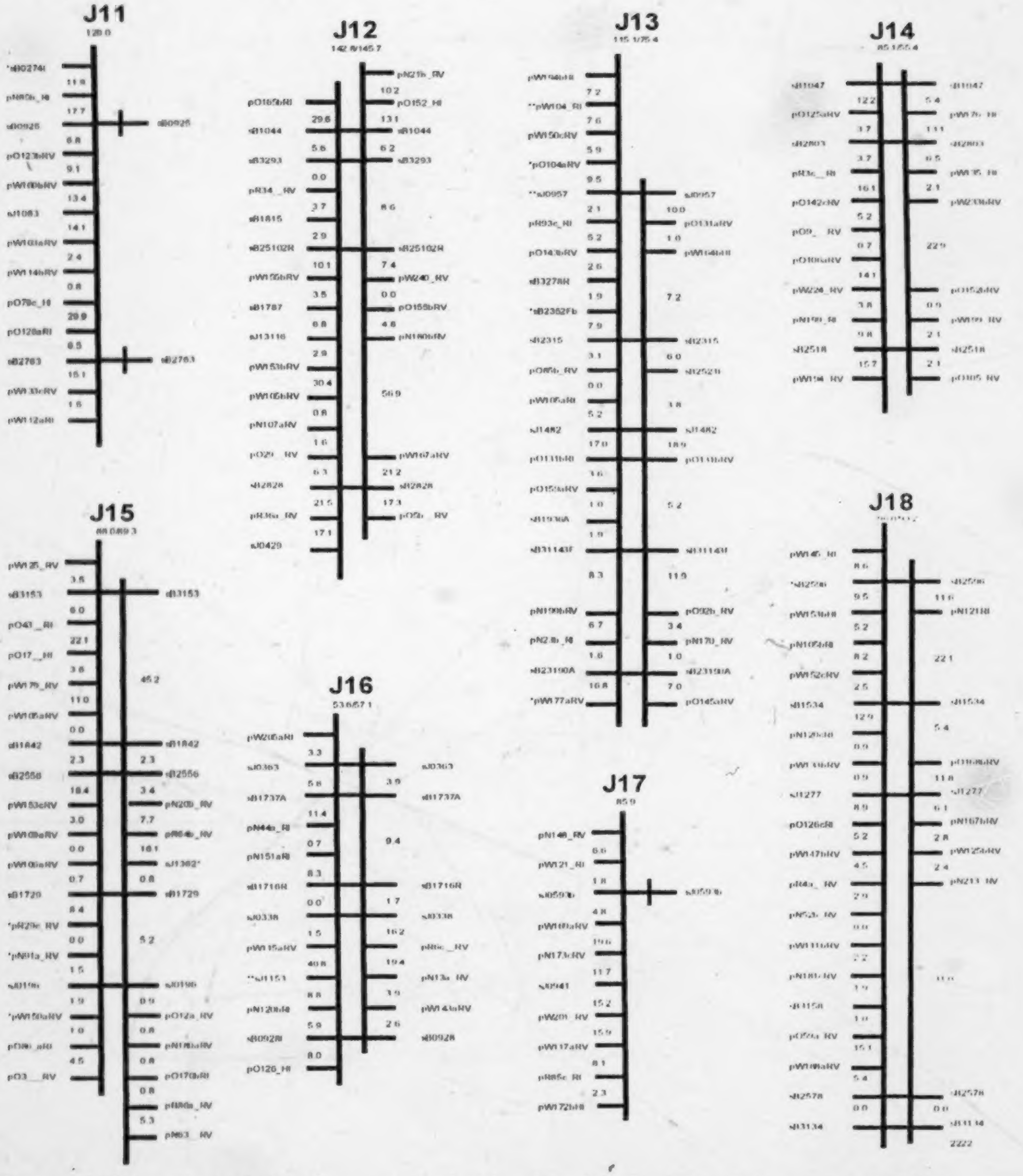
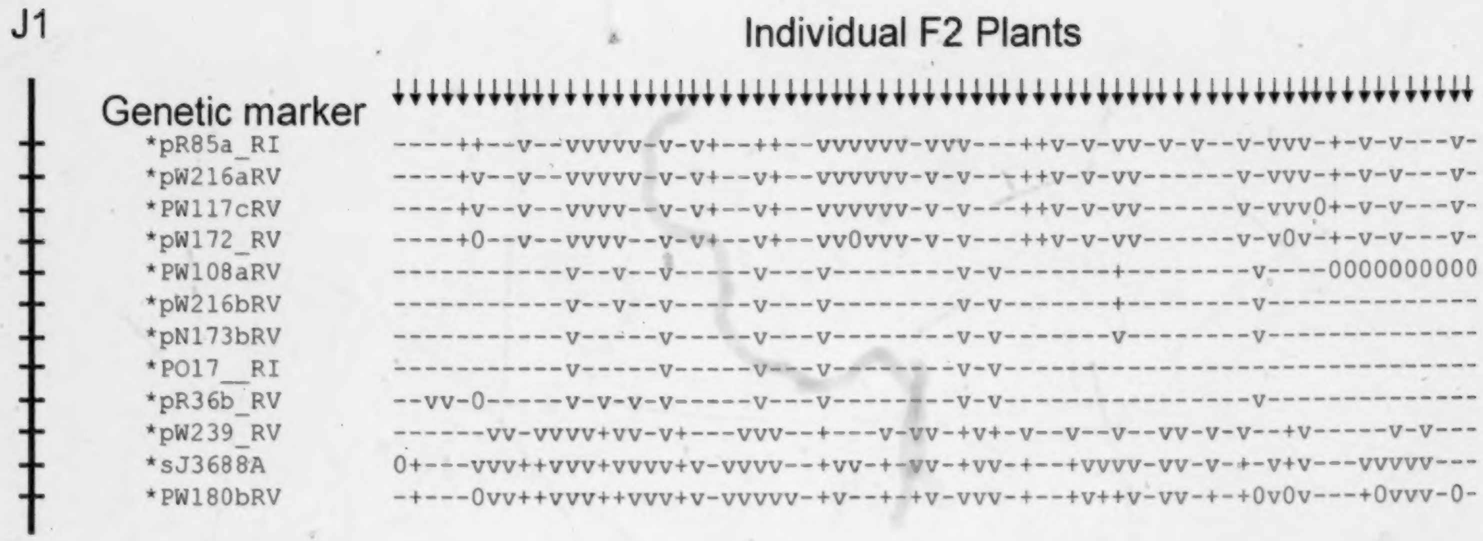


Figure 2-3 continued. Marker names that begin with the letter (p) represent DNA fragments used to probe Southern blots of genomic DNA to detect RFLP loci. Marker names that begin with the letter (s) represent primer pairs used to amplify specific microsatellite loci. Lower case letters immediately following RFLP probe names distinguish different loci detected by the same probe. Upper case letters following RFLP probe name indicate the restriction enzyme used to create the polymorphism, RV = Eco RV; RI = Eco RI; HI = Bam HI. Distances between markers in Kosambi centimorgans are listed in the intervals between loci, which are represented with even spacing. (*) indicate loci exhibiting segregation distortion, (*) = $p<0.05$, (**) = $p< 0.01$.

Allele segregation on linkage group J1 was severely distorted through the entire length, with the most extreme distortion in the region bounded by pO17RI and pR36bRV, where only one individual contained two loci homozygous for AC Vulcan alleles and only eight others displayed any heterozygosity in this region. Every other individual in this population contained genetic material only from UM3132 in this area (Figure 2-4).

2.3.3. Mapping the dominant gene for resistance to *L. maculans*, *LMJR1*

For each marker, the number of resistant F2 individuals homozygous or heterozygous for AC Vulcan alleles, plus those susceptible individuals homozygous for UM3132 alleles was tabulated. The number of susceptible F2 individuals homozygous or heterozygous for AC Vulcan alleles, plus those resistant individuals homozygous for UM3132 alleles was also tabulated. The ratio of these two classes was calculated and compared against an expected ratio assuming random assortment of alleles in a chi-squared goodness-of-fit test. The markers sB1936A, pO159aRV, sB31143F, pO131bRI, sJ1482, pW105aRI, pO85bRV, pO92bRV, and pN34cRV all from J13 were found to deviate significantly (p-value < 0.05) from the expected 1:2:1 ratio of alleles in resistant and susceptible classes. The marker that deviated most significantly was sB1936A (χ^2= 10.05, p-value= 0.0015, df= 1). Although there is a definite association between resistance to blackleg and markers in this region of J13, a precise location was difficult to determine. For the dominant gene for resistance, a specific genotype cannot be assigned for each phenotype since any resistant plant may be homozygous or heterozygous for the gene conferring resistance or may be resistant due to the action of the recessive gene for resistance. To more accurately determine the position of the dominant gene for resistance



```
Genetic marker
   *pR85a_RI     ----++--v--vvvvv-v-v+--++--vvvvvv-vvv---++v-v-vv-v-v--v-vvv-+-v-v---v-
   *pW216aRV     ----+v--v--vvvvv-v-v+--v+--vvvvvv-v-v---++v-v-vv------v-vvv-+-v-v---v-
   *PW117cRV     ----+v--v--vvvv--v-v+--v+--vvvvvv-v-v---++v-v-vv------v-vvv0+-v-v---v-
   *pW172_RV     ----+0--v--vvvv--v-v+--v+--vv0vvv-v-v---++v-v-vv------v-v0v-+-v-v---v-
   *PW108aRV     -----------v--v--v-----v---v--------v-v-------+--------v----0000000000
   *pW216bRV     -----------v--v--v-----v---v--------v-v-------+--------v--------------
   *pN173bRV     -----------v-----v-----v---v--------v-v-------v--------v--------------
   *PO17__RI     -----------v-----v-----v---v--------v-v-------------------------------
   *pR36b_RV     --vv-0-----v-v-v-v-----v---v--------v-v-----------------v--------------
   *pW239_RV     ------vv-vvvv+vv-v+---vvv--+---v-vv-+v+-v--v--v--vv-v-v--+v-----v-v---
   *sJ3688A      0+---vvv++vvv+vvvv+v-vvvv--+vv-+-vv-+vv-+--+vvvv-vv-v-+-v+v---vvvvv---
   *PW180bRV     -+---0vv++vvv++vvv+v-vvvvv-+v--+-+v-vvv-+--+v++v-vv-+-+0v0v---+0vvv-0-
```

Figure 2-4. Marker scoring data for J1. (+) indicates an individual is homozygous for the AC Vulcan allele for that marker. (v) indicates individual is heterozygous for that marker. (-) indicates that individual is homozygous for the UM3132 allele for that marker. Segregation is extremely distorted along the entire linkage group.

to blackleg, markers covering linkage group J13 were screened on the B1 population using UM3132 as the recurrent parent. In this population, the resistance trait, dubbed *LMJR1* mapped to an interval between pN199RV and sB31143F (Figure 2-5). The distances between *LMJR1* and the flanking markers are large (22.1cM from pN199RV, 8.7cM from sB31143F), but these may be inflated. There are 3 individuals out of 80 that have double crossovers flanking *LMJR1*. Selfed progeny from these individuals were tested and found to confirm the initial phenotypic scoring. In each case the individual scored susceptible when the flanking markers indicated the presence of *LMJR1*. It is possible that there are other, redundant background genetic factors which are present in AC Vulcan, but not UM3132, which are required for the proper functioning of *LMJR1* and that these are rare individuals where none of the redundant factors are present.

2.3.4. Predicting a probable location for a recessive resistance gene to *L. maculans*

Having determined a location for the dominant gene conferring resistance, *LMJR1*, it became possible to infer a genotype for the recessive gene for resistance for a subset of 31 informative individuals from both subpopulations. Individuals could be identified which were resistant but did not appear to contain *LMJR1* based on data from the flanking markers, implying that they must be homozygous for the recessive resistance gene. Segregation data from F3 progeny could also be used to infer a genotype for the F2 parent. Using the genotype for this subset, a probable location for the second resistance gene, between the microsattelite markers sB1534 and sJ1277 on linkage group J18 was identified. The inferred genotype of the recessive resistance gene shows linkage to marker sB1534 at LOD 2.51 and a distance of 24.05 Kosambi cM. Given the relatively small size of the informative subset, the limited number of markers spanning both subpopulations and the potential for misclassification of F2 phenotypes, this trait cannot currently be mapped more precisely.

2.4. Discussion

Previous studies of resistance in *B. juncea* suggested models of inheritance that consisted of anywhere from one to three genes controlling the response to *L. maculans* (Pang and Halloran 1996; Chevre et al. 1997; Keri et al. 1991). The work presented here saw a segregation ratio of 144: 34 resistant to susceptible plants in a natural F2 population. These results are compatible with either a one gene model or a two gene model with one dominant and one recessive resistance gene. F3 families were tested to distinguish between the two possibilities. The presence of susceptible F2 individual derived F3 families which segregate for resistance consistent with a ratio of 1 resistant: 3 susceptible plants rule out the possibility of a single dominant gene conferring resistance to blackleg in this population. These results are in agreement with the proposed two gene model suggested by Keri et al. (1991).

Phenotypic scoring on all plants in this study was based on single plant ratings at the cotyledonary stage. Since *L. maculans* is quite capable of growing saporophytically, premature senescence of cotyledons could plausibly inflate disease ratings and make resistant plants appear to be susceptible. Conversely, disease escapes could potentially be produced when innoculum droplets failed to adhere to the leaf surface. In either case, given the lack of replication these ratings are prone to error. In cases where the initial rating for a particular plant was suspected to be an error, attempts were made to examine

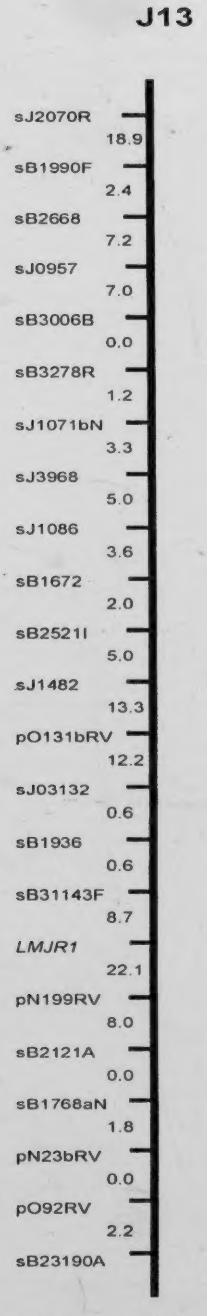

J13
sJ2070R
18.9
sB1990F
2.4
sB2668
7.2
sJ0957
7.0
sB3006B
0.0
sB3278R
1.2
sJ1071bN
3.3
sJ3968
5.0
sJ1086
3.6
sB1672
2.0
sB2521I
5.0
sJ1482
13.3
pO131bRV
12.2
sJ03132
0.6
sB1936
0.6
sB31143F
8.7
LMJR1
22.1
pN199RV
8.0
sB2121A
0.0
sB1768aN
1.8
pN23bRV
0.0
pO92RV
2.2
sB23190A

Figure 2-5. Linkage group J13, generated from B1 segregation data. Marker names that begin with (p) represent DNA fragments used to probe Southern blots of genomic DNA to identify RFLP loci. Marker names that begin with (s) represent primer pairs used to amplify specific microsatellite loci. Markers separated by recombination are shown at evenly spaced intervals. The map distance between markers in Kosambi centimorgans is shown in the intervals between markers. The position of *LMJR1* is shown between pN199RV and sB31143F.

that plant's offspring to clarify the rating. In the majority, these tended to reveal false susceptible scores. All susceptible individuals that could be tested in the F2 generation were tested.

The genetic map presented here shares common markers with previously published maps of various *Brassica* species. A comparison of this map and the *B. juncea* map published by Axelsson et al. (2000), shows that they are very similar with respect to marker order and groupings with a few exceptions. A total of six marker pairs are in inverted orders between the two maps on linkage groups J3, J6, J7, J9, J15, and J18. Comparison with the map described by Axelsson et al. (2000), allowed the identification of the A and B genomes.

Marker scorings for the central region of J1 exhibited extreme segregation distortion with almost no contributions in this area from AC Vulcan across the population. Segregation distortions this severe are very rare, especially in a natural cross. A genetic map in *Lablab purpureus* exhibited loci with this same level of extreme distortion (Konduri et al. 2000). It is possible that these loci are linked to a cuckoo gene, similar to cuckoo chromosomes in various grass species which have the ability to induce lethal mutations and chromosome breakages in gametes that do not possess them (Endo 1991).

The dominant gene conferring resistance, dubbed *LMJR1* was located on linkage group J13. This linkage group is part of the *B. juncea* B genome, and corresponds to linkage group G3 in *B. nigra* (Lagercrantz and Lydiate 1995). Interestingly, a cDNA derived from *B. nigra* that confers increased resistance when transformed into susceptible

B. napus, denoted *Lm1*, was shown to map on linkage group G3 in *B. nigra*, only 1.2 cM away from marker pO131 (Wretblad et al. 2003). *LMJR1*, described here maps 22.1 cM distant from the equivalent locus. Although unlikely, it is not impossible given the differences in mapping populations and species used, that *Lm1* and *LMJR1* are at the same locus. However, *Lm1* also maps on G8 which has also previously been associated with resistance to blackleg (Dixelius and Wahlberg 1999; Dixelius 1999). It is common for RFLP markers in *Brassica* spp. to detect multiple loci. Thus *Lm1*, used as an RFLP marker may be detecting related sequences on G3 and G8. However, *Lm1* was isolated using a PCR approach and degenerate primers designed from the *Pto* gene of tomato and the *Hm1* gene from Maize (Wretblad et al. 2003). Analysis of the full cDNA suggests that it may be chimeric, and its position on both G3 and G8 may be an artifact. In fact the cDNA has significant sequence similarity to two unrelated *A. thaliana* genes. The 5' end of the cDNA closely matches a sequence from *A. thaliana* chromosome 5, while the 3' end has homology to sequences from *A. thaliana* chromosomes 1 and 4. There is no overlap between the 5' and 3' regions of homology. To elucidate the possibility of a relationship between *Lm1* and *LMJR1*, it would be useful to use this cDNA as a probe on the populations described here.

Plieske et al. (1998) investigated *L. maculans* resistance in *B. napus*/B genome hybrid lines. Markers associated with introgressed traits for resistance were mapped in the *B. nigra* population used by the group of Dixelius, they were also located on G3 and linked to the *B. nigra* cDNA (Plieske and Struss 2001; Wretblad et al. 2003). Combined, these results strongly suggest that this region of J13/G3 is the site of at least one and possibly multiple genes controlling resistance.

There are only a few published reports of mapped genes for resistance to *L. maculans* in *B. napus*. Two of these (Ferreria et al. 1995; Mayerhofer et al. 1997) suggest that these resistance genes are both on *B. napus* linkage group 6. This linkage group 6, although developed using a different marker set, has been demonstrated to correspond to *B. napus* linkage group N7 on the map of Parkin et al. (1995). *B. napus* linkage group N7 is part of the A genome from the diploid species *B. rapa*. There is some homology between N7 and part of G3/J13 (Lagercrantz and Lydiate 1996). However, the segment of J13/G3 containing the resistance gene is homologous to N8/N18 in *B. napus* indicating that there is likely no relation between the described *B. napus* resistance genes and *LMJR1*.

Although, this population is segregating for two genes conferring resistance, only the dominant gene could effectively be mapped. A probable location for the recessive resistance gene has been identified. However, the use of an F2 population segregating for two resistance genes reduced the amount of information useable for mapping, making a confident and precise determination of the position of this second gene difficult. It will be necessary to develop a population segregating only for the recessive gene conferring resistance to definitively locate this trait. There are numerous ways such a population could be created; the simplest would simply be to use the segregating F3 progeny from an F2 plant susceptible to blackleg.

In summary, a detailed linkage map was constructed from a population of *B. juncea* segregating for blackleg resistance. It was possible to locate a dominant resistance trait, *LMJR1* on this map. This trait likely represents a novel source of blackleg resistance and is currently the only gene conferring resistance to blackleg to be mapped in a natural population of any *Brassica* B genome containing species. Further study of the materials described here will allow the study of a second trait for resistance from this source.

Works Cited

Axelsson T., Bowman C.M., Sharpe A.G., Lydiate D.J., Lagercrantz U. (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43: 679-688.

Chevre A.M., Barret P., Eber F., Dupuy P., Brun H. Tanguy X., Renard M. (1997) Selection of stable *Brassica napus-B. juncea* recombinant lines resistant to blackleg (*Leptosphaeria maculans*). 1. Identification of molecular markers, chromosomal and genomic origin of the introgression. Theor Appl Genet 95: 1104-1111.

Dixelius C. (1999) Inheritance of the resistance to *Leptosphaeria maculans* of *Brassica nigra* and *B. juncea* in near-isogenic lines of *B. napus*. Plant Breeding 118: 151-156.

Dixelius C., Wahlberg, S. (1999) Resistance to *Leptosphaeria maculans* is conserved in a specific region of the *Brassica* B-genome. Theor Appl Genet 99: 368-372.

Delourme R., Foisset N., Horvais R., Barret P., Champagne G., Cheung W. Y., Landry B. S., Renard, M. (1998) Characterisation of the radish introgression carrying the *Rfo* restorer gene for the *Ogu*-INRA cytoplasmic male sterility in rapeseed (*Brassica napus* L.). Theor Appl Genet 97: 129-134.

Delwiche P.A. (1980) Genetic aspects of blackleg (*Leptosphaeria maculans*) resistance in rapeseed (*Brassica napus*). Ph.D. Thesis, University of Wisconsin. Madison. WI.

Endo T.R. (1991) Gametocidal chromosomes and their induction of chromosome mutations in wheat. Jpn. J. Genet. 65: 135-152.

Ferreria M.E., Rimmer S.R., Williams P.H. Osborn T.C. (1995) Mapping Loci controlling *Brassica napus* resistance to *Leptosphaeria maculans* under different screening conditions. Phytopathology 85: 213-217.

Keri M. (1991) Resistance of *Brassica juncea* Czern & Coss to blackleg disease caused by *Leptosphaeria maculans* (Desm.) Ces. Et de Not. M.Sc. thesis. University of Manitoba, Winnipeg, Manitoba.

Keri M., van den Berg C.J.G., McVetty P.B.E. Rimmer S.R. (1997) Inheritance of resistance to *Leptosphaeria maculans* in *Brassica juncea*. Phytopathology 87: 594-598.

Konduri V., Godwin I.D., Liu C.J. (2000) Genetic mapping of the *Lablab purpureus* genome suggests the presence of 'cuckoo' gene(s) in this species. Theor Appl Genet. 100: 866-871.

Kosambi D.D. (1944) The estimation of map distances from recombination values. Ann. Eugen. 12: 172-175.

Lagercrantz U, Lydiate D..J. (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. Genome 38: 255-264.

Lagercrantz U., Lydiate D.J. (1996) Comparitive genome mapping in Brassica. Genetics 144: 1903-1910.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E. Newberg L. (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Mayerhofer R., Bansal V.K., Thiagarajah M.R., Stringam G.R., Good A.G. (1997) Molecular mapping of resistance to *Leptosphaeria maculans* in Austrailian cultivars of *Brassica napus*. Genome 40: 294-301.

Pang E.C.K. Halloran G.M. (1996) The genetics of adult-plant blackleg (*Leptosphaeria maculans*) resistance from *Brassica juncea* in *B. napus*. Theor Appl Genet 92: 382-387.

Parkin I.A.P., Sharpe A.G., Keith D.J., Lydiate D.J. (1995) Identification of the A and C genomes of Amphidiploid *Brassica napus* (oilseed rape). Genome 38: 1122-1131.

Plieske J., Struss D. (2001) STS markers linked to *Phoma* resistance genes of the *Brassica* B-genome revealed sequence homology between *Brassica nigra* and *Brassica napus*. Theor Appl Genet. 102: 483-488.

Plieske J., Struss D., Robbelen G. (1998) Inheritance of resistance derived from the B-genome of *Brassica* against *Phoma lingam* in rapeseed and the development of molecular markers. Theor Appl Genet 97: 929-936.

Rimmer S.R., van den Berg C.G.J. (1992) Resistance of oilseed *Brassica* spp. To blackleg caused by *Leptosphaeria maculans*. Can J Plant Path 14: 56-66.

Roy N.N. (1978) A study on disease variation in the population of an interspecific cross of *Brassica juncea* L. x *B.napus* L. Euphytica 27: 145-149.

Roy N.N. (1984) Interspecific transfer of *Brassica juncea* - type high blackleg resistance to *Brassica napus*. Euphytica 33: 295-303.

Sharpe A.G., Parkin I.A.P., Keith D.J., Lydiate D.J. (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38: 1112-1121.

Uzonova M., Ecke W., Weissleder K., Robbelen G. (1995) Mapping the genome of rapeseed (*Brassica napus* L.) I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90: 194-204.

Wretblad S., Bohman S., Dixelius C. (2003) Overexpression of a *Brassica nigra* cDNA gives enhanced resistance to *Leptosphaeria maculans* in *B. napus*. Mol. Plant-Microbe Interact. 16: 477-484.

3. Analysis of EST libraries from *Leptosphaeria maculans* and infected *Brassica napus* leaf tissue

3.1. Introduction

Both plant resistance to pathogens and fungal virulence on hosts can be inherited as simple, single genetic traits exhibiting a gene for gene interaction in virtually all interactions. However, an examination of the underlying mechanisms of resistance and virulence reveals a complex battle between the two organisms. This struggle involves recognition of the other partner, activation of numerous signaling networks to trigger responses and the mounting of multiple offenses and defenses. While an analysis and understanding of the interactions between single resistance genes and their avirulence cognates is incredibly important and especially useful for breeding for resistance, it does not go far in illuminating many of the questions relating to the biological processes involved in infection or resistance.

Although a general model has been established for the cellular response to pathogen attack in plants, none of the specific genes involved in the recognition of *Leptosphaeria maculans* by *Brassica* species have been described in molecular detail. Also, although the number of pathogen genes and molecular processes known is increasing rapidly, only two pathogenicity genes have been discovered for *L. maculans* (Idnurm and Howlett 2002; Idnurm and Howlett 2003). In fact, only 167 sequences from *L. maculans* have been deposited in Genbank. Of these, only 16 are complete coding sequences from identified genes, the rest are EST sequences (http://www.ncbi.nlm.nih.gov/). In order to gather sequence data for *L. maculans* as well as to gain a sample of the genes expressed by both organisms in an infection, EST libraries were created.

Collections of ESTs provide a relatively inexpensive way to gain valuable sequence data from organisms that may not yet be well represented in sequence databases (Pandey and Lewitter 1999). EST collections also have some advantages over genomic sequencing for particular applications. Because the process is based on an initial mRNA extraction step, only coding sequences are collected. In addition, EST libraries can be constructed so that they are tissue specific and can be enriched for sequences that are expressed only under certain conditions. This enables EST libraries to be much more efficient than genomic sequencing at addressing specific questions and also to provide very limited data on expression patterns. Because of this, EST libraries are often used as resources for gene discovery. They provide a relatively simple way of sampling large numbers of genes expressed at specific times in specific tissues or under specific environmental conditions.

EST library approaches have been used before to study plant pathogen interactions. Libraries have been made from soybean tissues infected with *Phytophthora sojae* (Qutob et al. 2000, Hraber and Weller, 2001); rice tissue infected with *Magnaporthe grisea* (Soonok et al. 2001); potato tissue infected with *Phytophthora infestans* (Ronning et al. 2003) and one other library consisting of sequence tags collected from *Brassica napus* infected with *L. maculans* (Fristensky et al, 1999). In the first two cases compatible reactions between plant and pathogen were used to maximize the recovery of ESTs from the pathogen. In the second two cases the authors chose incompatible reactions to focus on plant resistance responses. Accordingly the proportion of pathogen derived transcripts was estimated 75% in the *P.sojae*/soybean interaction (Qutob et al. 2000, Hraber and

Weller, 2001), 13 % of the ESTs with significant BLAST matches in the *M. grisea*/rice interaction (Soonok et al. 2001), 4.6% in the *P. infestans*/potato incompatible reaction (Ronning et al. 2003) and less than 1% in the previous *L. maculans/B. napus* interaction (Fristensky et al, 1999).

One of the goals of this research was to increase the availability of *L. maculans* sequences and gain understanding of the molecular biology and pathogenicity of this organism. ESTs were collected from a compatible leaf reaction after eight days of progress as well as *in vitro* cultured fungus. Eight days post infection was chosen as at this time colonization of plant tissues has become extensive and disease symptoms begin to be apparent, indicating that the fungus has switched to a more aggressive type of growth and that plant responses are being mounted (Hammond and Lewis, 1987; Oliver et al. 1993).The other goal of this research was to identify host genes involved in defense against pathogen attack. Despite using a susceptible host for the infection library, it was expected that defense genes would be expressed. Previously a chitinase isolated from *B. napus* was shown to be expressed in both resistant and susceptible cultivars in response to inoculation with *L. maculans* (Rasmussen et al. 1992a). Expression was initially delayed in the susceptible cultivar, but by eight days post-infection expression levels were roughly the same between the two cultivars (Rasmussen et al. 1992b).

Presently, 741 ESTs from *in vitro* cultured *L. maculans* and 1281 ESTs from infected leaf tissue have been collected. All sequences have been deposited in a searchable database complete with their top five nucleotide and predicted protein BLAST matches. Genes with putative roles in pathogenicity and resistance have been identified. This collection also serves as the foundation for microarray gene expression studies and for screening fungal knockout libraries to test gene function.

3.2. Methods and Materials

3.2.1. Culture and growth conditions

L. maculans PG2 strain PL86-12 from the collection at AAFC Saskatoon Research Centre was grown on V-8 agar (Chapter 2). A 1cm^2 plug of agar containing actively growing mycelium was transferred to 200ml liquid V-8 media. This culture was grown for 5 days at room temperature with shaking.

Seeds of *B. napus* accession N-0-1 (a doubled haploid line derived from the cultivar Westar and susceptible to PG2 isolates of *L. maculans*) were grown in a growth cabinet under standard growing conditions. Plants were grown for approximately four weeks and mature leaves were chosen for inoculation. Inoculations were performed with a 10^7 spores/ml suspension of PL86-12 pycnidiospores prepared as described in chapter two. Spores were infiltrated through the bottom leaf surface using a 10ml syringe without needle. The syringe tip was held flat against the underside of the leaf surface and spore suspension was gently forced into the leaf until approximately 75% of the leaf area showed signs of water soak.

3.2.2. RNA extraction

Five grams of 5 day old fungal mycelia and 10g each of 8 day infected and non-infected leaf tissues were flash frozen in liquid N_2 and ground to a fine powder. Total RNA was extracted using Amersham's RNA extraction kit following manufacturers protocols. mRNA was isolated using Amersham's mRNA purification kit following manufacturers instructions.

3.2.3. cDNA library construction

Three separate cDNA libraries were constructed. One from *in vitro* cultured *L. maculans*, one from infected *B. napus* leaf tissue, and one from uninfected *B. napus* leaf tissue. For each library 5μg of mRNA was used for 1st strand cDNA synthesis. Both the cDNA sysnthesis and library cloning were accomplished using the Superscript Plasmid System for cDNA Synthesis and Plasmid cloning kit (Gibco BRL Life Technologies). The cDNAs were directionally cloned into Sal I/ Not I digested pSPORT1 or pSPORT2. These vectors are identical except for the orientation of their multiple cloning sites. The infected *B. napus* leaf tissue library was subtracted in the following manner. First, target ssDNA from the infected *B. napus* leaf tissue library was produced from the f1 origin of replication by adding M13K07 helper phage (Invitrogen Life Technologies) to a library culture grown in DH12S electrocompetent cells (Invitrogen Life Technologies). Driver RNA was produced by digesting 20μg of each of the *L. maculans* library and the *B. napus* uninfected leaf library dsDNA plasmids with Sal I. T7 RNA polymerase and 1mM each ATP, UTP and GTP, 0.5mM CTP, 2mM biotin-14-CTP were used to produce biotinylated RNA copies of insert sequences. Target ssDNA was blocked with the Not I oligo dT in a reaction with 20 units *Taq* polymerase and *Taq* buffer with 600μM dTTP. Blocked ssDNA Target and biotinylated driver RNA sequences were hybridized in 80%

formamide, 100mM HEPES, pH 7.5, 2mM EDTA, 0.2M NaCl and 0.2% SDS for 24 hrs. at 42°C. A ratio of 67:1 uninfected leaf to infected leaf sequences was used, as well as a ratio of 10:1 cultured *L. maculans* to infected leaf sequences. Unhybridized, biotinylated RNA and target/driver hybrids were extracted with three rounds of binding with 25μg streptavidin followed by Phe:Chl extraction. Unhybridized target ssDNA was precipitated in the presence of 5μg yeast tRNA and used for transformation with DH5aE competent cells (Invitrogen Life Technologies).

3.2.4. Plasmid preparation

cDNA library containing cultures were plated out on LB agar containing 50μg/ml carbenicillin and X-gal. White colonies were selected and transferred individually to 96 well blocks containing 1.4ml liquid TB media and 50mg/ml carbenicillin and incubated overnight at 37°C. Alkali lysis plasmid extraction was performed using the QIAprep 96 Turbo miniprep kits according to the manufacturer's protocol (Qiagen). Purified plasmids were quantified fluorometrically in the presence of hoescht dye and diluted to a final concentration of 100ng/ml.

3.2.5. cDNA sequencing

Sequencing reactions were carried out using in an ABI 877 integrated thermocycler using Big Dye v3.0 (Applied Biosystems) sequencing chemistry half reactions with 300ng template and 3.2pmol of either M13 Forward or M13 reverse universal primers. Cycle sequencing was accomplished with the following cycle conditions: 15 cycles of 96°C for 10s; 55° for 5s; 70°C for 1 min followed by 15 cycles of 96°C for 10s; 70°C for

1 min. with 1°C/s ramps between cycle steps. Finished sequencing products were purified to remove unincorporated nucleotides. This was done either with Bio-Rad SEQueaky clean dye terminator removal kit (Bio-Rad laboratories) or with a simple ethanol precipitation. Purified sequencing products were run out on a 96 lane gel in an ABI 377 DNA sequencer.

3.2.6. Sequence analysis

All EST sequences were trimmed for vector sequences and sequence quality using Sequencher software (Gene codes corporation, Ann Arbor, MI) before being aligned to identify overlapping forward and reverse sequences. All EST sequences were deposited in the customized Microsoft Access database LESTer (*Leptosphaeria* EST Search Tool) along with the top 5 BLAST results against the Genbank nr database for each of blastn and blastx algorithms, as well as blastn results against *Aspergillus nidulans* and *Neurospora crassa* genomic sequence for each EST.

3.2.7. % G+C content and hexamer frequency analysis

A simple script in PERL was created to count the nucleotides in each EST from a FASTA formatted file of sequences and return the results as an output file with the sequence name and a total count for each nucleotide. For the hexamer analysis, a program originally described by White *et al.* (1993) was used. This program was written in the C programming language and can be downloaded from: (ftp://ftp.tigr.org/pub/software/qc/). This program employs a sliding hexamer window which tabulates the frequency of each of the 4,096 possible 6bp sequences in two training sets from different organisms. The frequency of hexamers is tabulated for a test sequence and this result is compared with the two training sets using a log-likelihood ratio function to generate a dissimilarity value (t-value) for the test sequence. In this case, negative t-values result when a sequence is more plant like and positive t-values result when a sequence is more fungal like, with t=0 being a distinct threshold.

Following the procedures outlined in Hraber and Weller (2001), *B.napus* coding sequences excluding mitochondrial, plastid and non-coding tRNA and rRNA sequences were retrieved from NCBI's Genbank using Entrez. The Entrez search was limited to also exclude EST, STS, GSS, working draft and patented sequences. *L. maculans* sequences were obtained in the same manner, however, there are only 111 sequences fitting this description in Genbank, so these were used to supplement the EST sequences from the LM library. All sequences were screened for poly dT or dA tails. Sequences that were less than 300nt were not used in the analysis

3.3. Results

3.3.1. General characteristics

cDNAs from three different libraries were single pass sequenced from both forward and reverse M13 universal primers. Each individual EST is named according to the library from which it was derived; LM6 and LM9 ESTs from cultured fungus. ILU from infected leaf un-subtracted and ILS from infected leaf subtracted. These library designations are followed by an individual clone number and in cases where the forward and reverse sequences did not overlap, either an F or R to designate which end of the cDNA the sequence represents. The average sequence length is 563 bp. From the LM cultured *Leptosphaeria* library 741 ESTs are present in the database LESTer, these represent 551 cDNAs clones. In this library, 60.1% of the 741 sequences are represented only once. The remaining 39.9% of the library can be grouped into 100 contigs with an average of 2.95 sequences per contig. Overall, this library is 16.5% redundant. A total of 301 ESTs from the ILU unsubtracted infected leaf library are present in the database, these represent 244 cDNA clones sequenced. From the ILS subtracted infected leaf library, 1390 ESTs are in the database, representing 1203 cDNA clones. In this library, 70.4% of the 1390 sequences are represented only once. The remaining 29.6% of the library can be grouped into 122 different contigs with an average of 3.4 sequences per contig. Overall, this library is 20.8% redundant.

The subtraction method used to create the ILS library depends on hybridization of biotinylated RNA transcripts to the ssDNA target library and removal of duplex and unhybridized RNA sequences with avidin. Because of this dependence on insert transcripts from the driver library, any vectors lacking an insert in the target library will not be subtracted out and will functionally behave the same as sequences unique to the target library. As empty vectors can be detected using blue/white selection on agar plates containing Xgal and IPTG, this provides an easy way to estimate the extent of enrichment in a subtracted library. In the ILU library prior to subtraction, 11 of 667 colonies counted were blue, indicating that 1.6% of the library clones were vector without insert. After subtraction, 487 of 2521 colonies counted were blue; the proportion of clones without insert had increased to 19.3%. This suggests ~ 10 fold enrichment for unique sequences.

3.3.2. Determining sequence origin in mixed libraries

When analyzing sequences from mixed cDNA libraries like those from infected tissues, determining the organism of origin for any particular sequence becomes an obstacle. Ideally large amounts of sequence data from the study organisms or their close relatives would be available. In this case BLAST analysis of the ESTs should result in the identification of high degrees of similarity between a particular EST and sequence from one of the study organisms involved or a close relative. Fortunately the model plant *Arabidopsis thaliana* for which the entire genome sequence is available, is very closely related to *B. napus*. Accordingly, 80% of the sequences in the ILS library have a high degree of nucleotide sequence similarity to *A. thaliana* or other *Brassica* species with an e-value of $1e^{-5}$ or less. Only 4% of the LM library has nucleotide sequence similarity to *A. thaliana* or *Brassicas* at e-value $1e^{-5}$ or less (Figure 3-1A). Although complete genome sequences for the filamentous fungi *Aspergillus nidulans* and *Neurospora crassa* are available, only 16% of the LM ESTs have any significant similarity to previously obtained fungal sequences (Figure 3-1A). Despite a general lack of strong similarity to other fungi at a nucleotide sequence level, *L. maculans* does exhibit strong similarities at a translated protein level, especially with *N. crassa*. Of the LM library ESTs, 57% have six frame translated EST similarities to filamentous fungi protein sequences with an e-value of $1e^{-5}$ or lower. Only 1% of LM sequences have similarity to plant sequences at $1e^{-5}$ or lower. Translated sequences having weak similarity to those present in the public database account for 35% of the LM library and the remaining 4% have no similarity

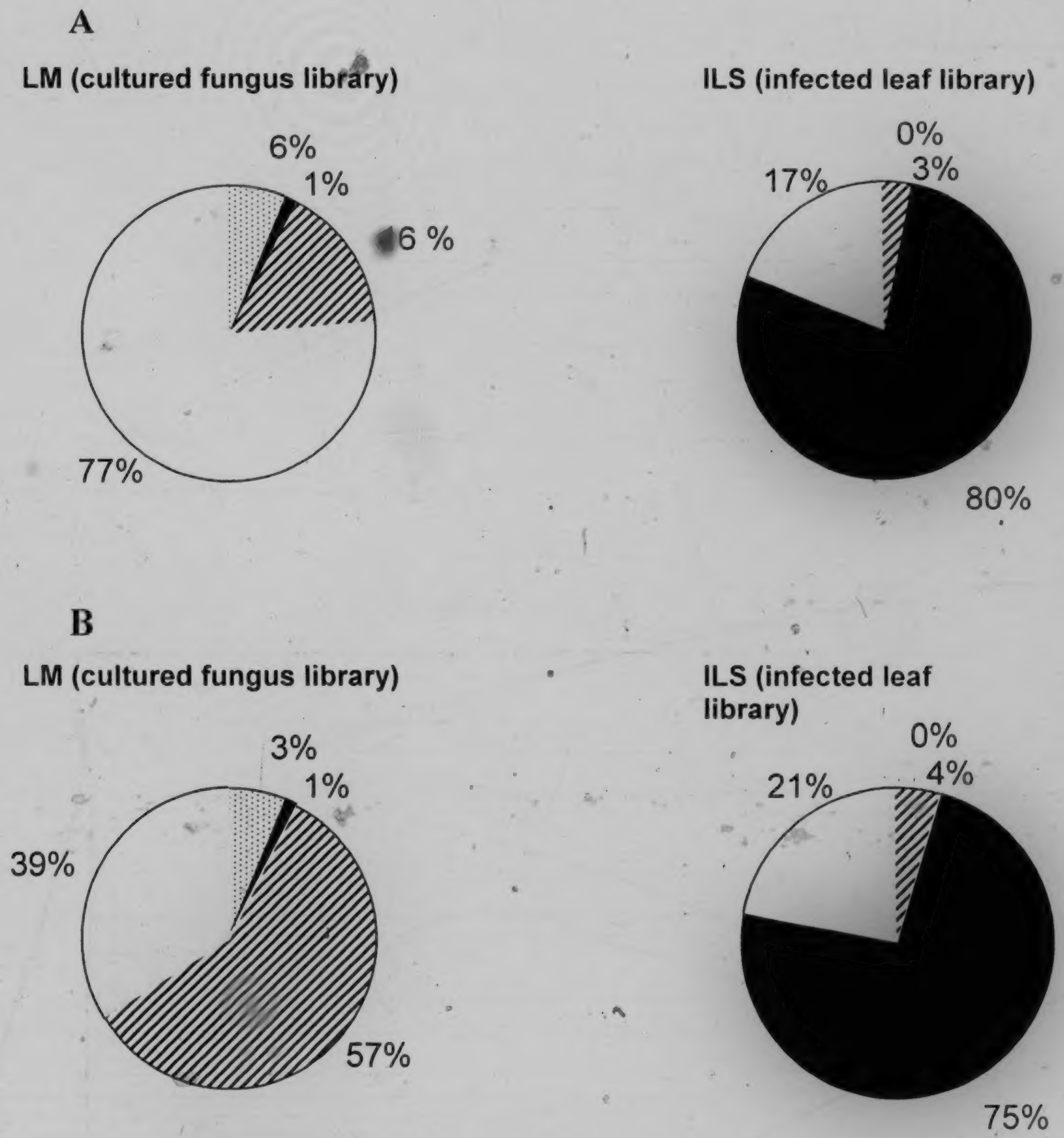


Figure 3-1. Similarity of library ESTs to sequences derived from other organisms. (A) results based on nucleotide sequences compared using the blastn algorithm. (B) results based on translated protein sequence comparisons using the blastx algorithm.

░ Animal (human, mouse, fruit fly, worm etc.)
■ *Arabidopsis* or other plants
▨ Fungi
□ Non-significant matches/unknown

whatsoever to anything previously sequenced (Figure 3-1B). In the ILS library, searches with translated ESTs reveal that 75% of the library has a significant similarity ($1e^{-5}$ or less) to *A. thaliana* or *Brassicas* and 4% to have significant similarity to fungi. 16% of the sequences have weak similarity to any record in genbank and 5% of sequences have no similarity whatsoever to any publicly available sequence in genbank (Figure 3-1B). Combining the results from both nucleotide and translated protein sequence comparisons. 57% of the LM library can be demonstrated to be very similar to known fungal sequences, a further 7% of the library appears to be similar to other organisms and the last 37% of the library does not look similar to anything previously sequenced either at the nucleotide or translated protein level. In the ILS library, combining both the nucleotide and translated protein BLAST results indicates that 82% of sequences are similar to plants, 5% sequences are similar to fungi and 13% cannot be classified using sequence similarity. Given that only slightly over half of the sequences in the LM library, which are by definition fungal in origin, could be identified as such using sequence similarity, it is not unreasonable to assume the true number of fungal sequences present in the ILS library may approach 10%. Identifying which particular sequences from this 'unknown' group are fungal still presents a problem.

In an effort to clarify the organism of origin for each particular sequence, other analysis techniques were undertaken. These included looking at %G+C content and at hexamer frequency. A sample of 426 *B. napus* coding sequences containing a total of 603,316 nucleotides was retrieved from Genbank using the Entrez batch retrieval tool. Also retrieved were 111 *L. maculans* coding sequences, these were added to the sequences from the LM library for a total of 788 sequences and 480,267 nucleotides. The percentage G+C content was calculated for each sequence in both sets. The *B. napus* sequences have a mean G+C content of 43%, with a standard deviation of 5.11, while the *L. maculans* sequences have a mean G+C content of 53% with a standard deviation of 4.91. These means are significantly different when compared using a students two tailed t-test assuming unequal variances (t= 31.5, P<0.0005). However, the G+C content distributions for each set overlap to an extent (Figure 3-2). The %G+C distribution for the ILS mixed library ESTs follows that of the *B. napus* sequence set very closely except for a slight increase in frequency of ESTs with G+C > 51%. This slight increase could correspond to the estimated 10% of sequences which are fungal in origin from the sequence similarity analysis above. In fact when the G+C content of each of the ILS ESTs is broken down by BLAST similarity to plant or fungal sequences, the putative fungal sequences all cluster to the upper range of G+C contents, approximately 46%-59% consistent with the range of *L. maculans* sequences (Figure 3-3). However, putative plant sequences are also common in this range and unknown sequences are scattered across the entire range of G+C contents (24%-59%). Thus it is not possible to definitively categorize these unknown sequences either through sequence similarity or %G+C content.

Hexamer word count frequencies have previously been used to determine sequence origin in mixed libraries (Hraber and Weller, 2001). Calibration curves were generated from the *B. napus* and *L. maculans* training sets used in the %G+C tests (Figure 3-4). The median value for each distribution is easily spotted from these curves as it is the value where the cumulative probability is equal to 0.5. The slope of the curve is inversely proportional to variance, and error rates can be determined by noting where

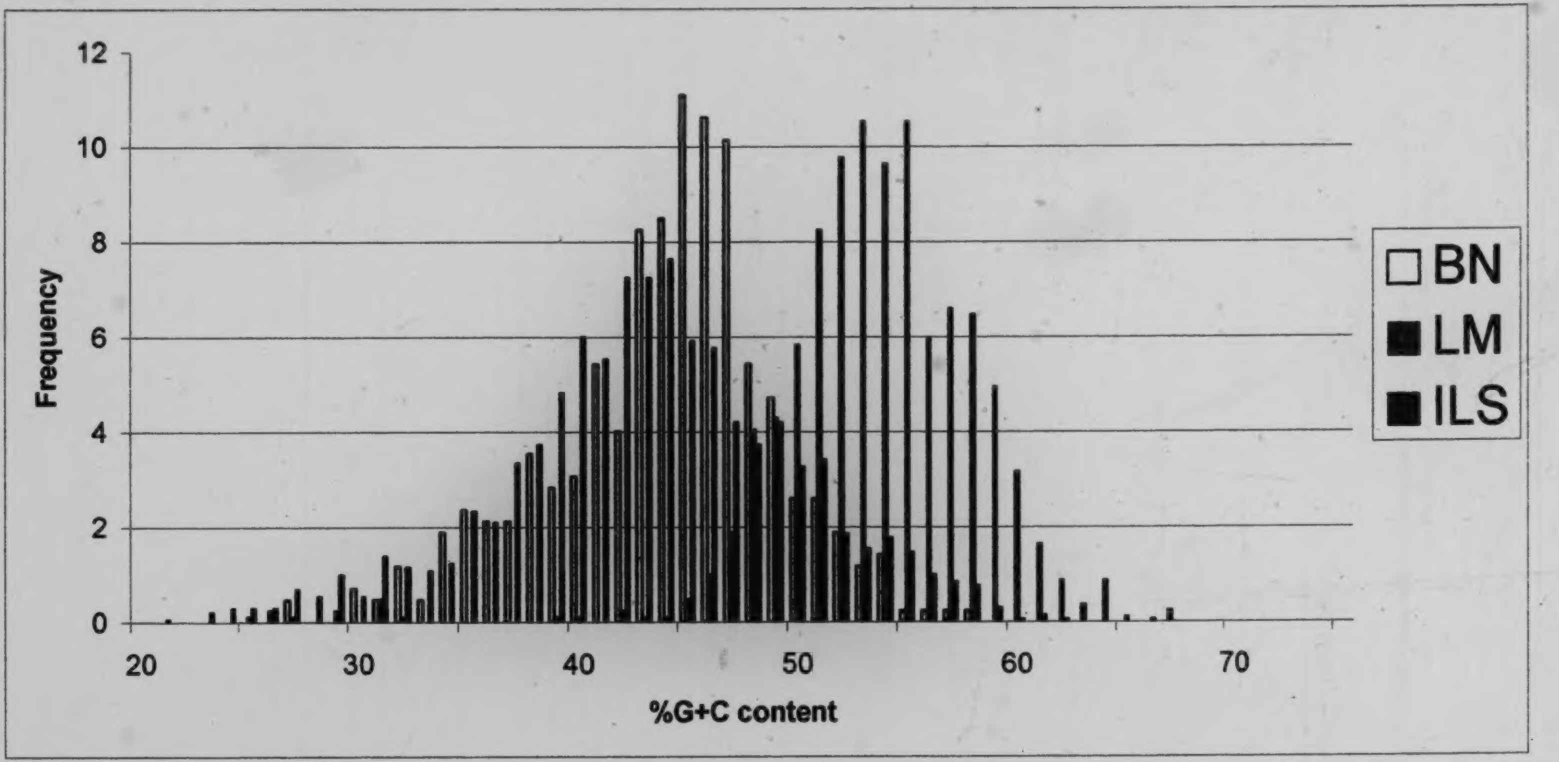


Figure 3-2. G+C content distribution for sequences from *B. napus* retrieved from genbank (BN), *L.maculans* from the library described in this chapter (LM) and the ILS library ESTs (ILS).

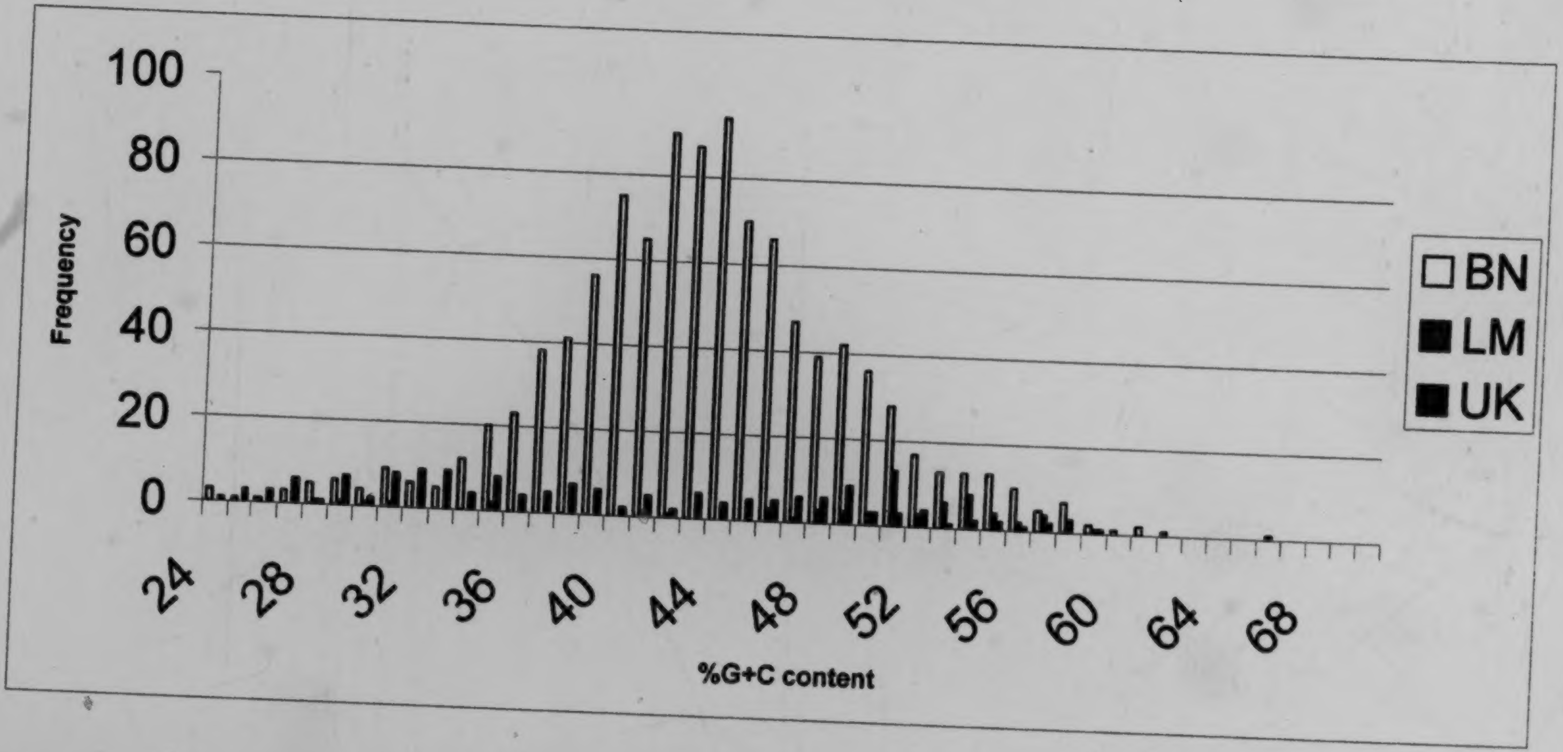


Figure 3-3. G+C content of the ILS library ESTs broken down by organism classification. BN, ESTs that are putatively derived from *B. napus* based on BLAST results.LM, ESTs that are putatively derived from *L. maculans* based on BLAST results UK, ESTs that could not be classified based on BLAST results

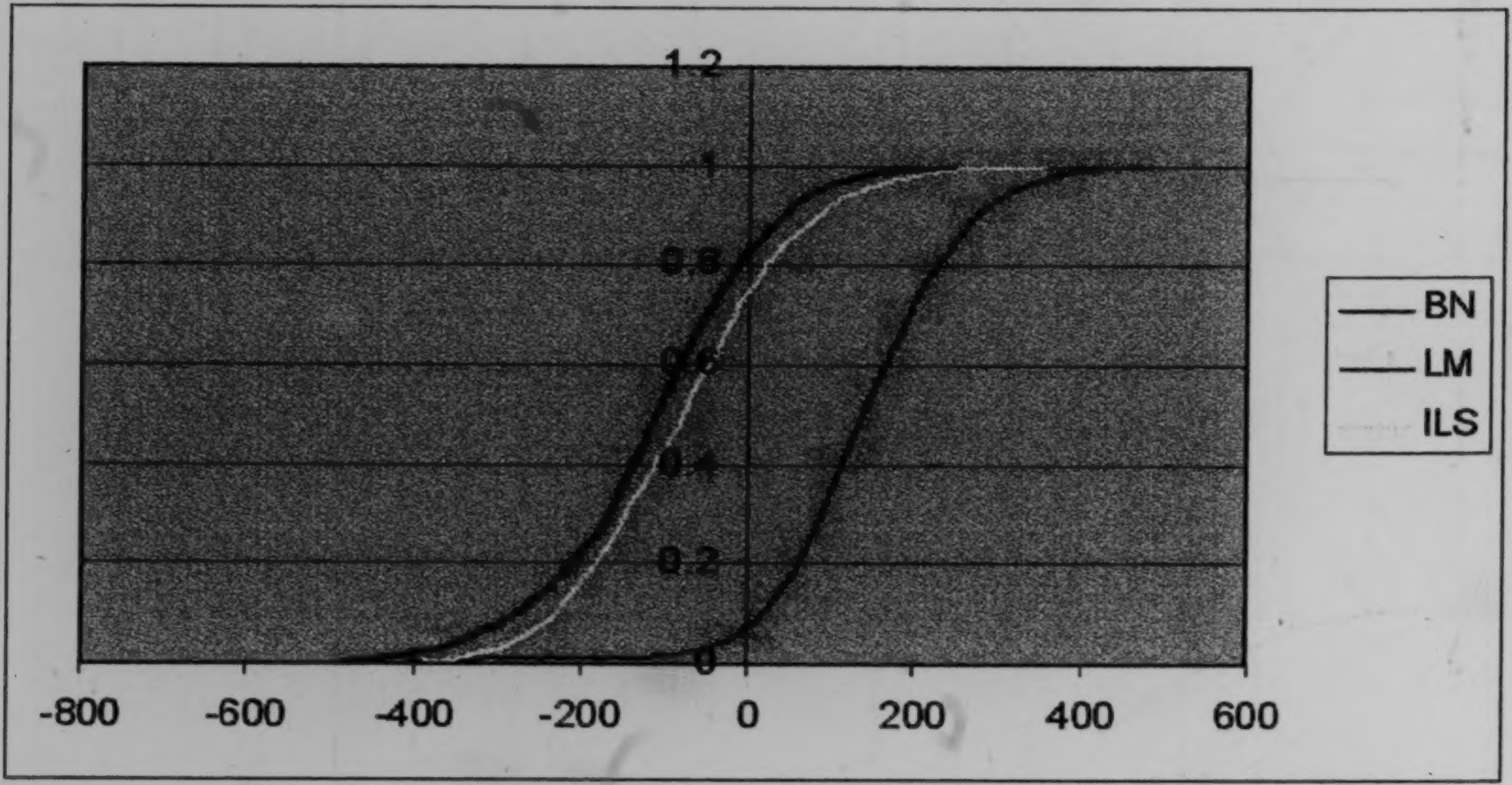


Figure 3-4. Hexamer composition analysis t-values. Shown are BN and LM calibration curves generated from testing random sequence fragments from the genbank retrieved *B.napus* or LM library *L. maculans* data sets against the hexamer frequency distributions for each entire data set. The ILS curve shows the ordered t-values for each ILS EST.

each calibration curve crosses t=0. When testing whether or not a given sequence is from *B. napus* the false positive and false negative rates can be read directly from these curves where they cross 0 (a= 0.170, b=0.074). Figure 3-4 shows the results of plotting t-values for the ILS EST library test sequences against the training sets. 25.0% of the ILS ESTs appear to be fungal in origin. Although, given the very high rate of false positives with this sample, most of these may be misidentified *B. napus* sequences. Comparing the result of this hexamer word count test with the results from sequence similarity searches, 79% of the ESTs identified as *B. napus* in origin based on sequence similarity were more plant like based on their hexamer composition. This is in line with the error rate predicted from the calibration curves. Of the putative *L. maculans* ESTs, 4% are misidentified, which is also in keeping with the expected false negative rate of 7.4%. Of the unknown ESTs, 63% appear to be more plant like than fungal, although due to the high false positive rate, this likely underestimates the true number of plant derived sequences in this category.

3.3.3. Analysis of putative functions

In the LM EST library 82 ESTs have similarity to hypothetical proteins with no annotated function. This leaves 374 ESTs for which a putative function may be assigned on the basis of sequence similarity. A list of ESTs with a possible role in pathogenesis is presented in Table 3-1. For the ILS library, 70 ESTs have similarity to hypothetical proteins with no annotated function leaving 1042 ESTs for which some inferences about function may be made. A list of ESTs that have a putative role in either defense or pathogenesis is presented in Table 3-2. There are four non-redundant ESTs in this library

Table 3-1. A list of ESTs of interest from the LM library, presented here are candidates for pathogenicity genes with descriptions of their closest translated protein similarity matches using the blastx algorithm.

Clone name	Description of sequences with closest similarity	Score	e-value
	Detoxification of plant compounds		
LM9-194R	(AF192405) putative cyanide hydratase [Leptosphaeria maculans]	365	1e-100
LM6-220R	(U62933) multidrug resistance protein 1 [Aspergillus fumigatus] gb|AAB88658.1|	244	6e-64
LM6-220F	multidrug resistance protein 1 - fungus (Filobasidium floriforme) gb|AAC49889.1|	187	6e-47
LM6-127F	multidrug resistance protein 1 - fungus (Filobasidium floriforme) gb|AAC49889.1|	135	2e-31
LM6-147F	Catalase B emb|CAA69069.1| (Y07763) catalase [Aspergillus fumigatus]	99.8	1e-20
LM6-127R	(U62933) multidrug resistance protein 1 [Aspergillus fumigatus] gb|AAB88658.1|	72.8	2e-12
LM6-147R	(AF327335) catalase [Blumeria graminis]	188	3e-47
LM9-357	(AF062654) manganese superoxide dismutase [Emericella nidulans] gb|AAK17008.1|	176	1e-82
	Metabolism		
LM6-5R	(NC_003424) glutamine synthetase [Schizosaccharomyces pombe] sp|Q09179|	339	2e-92
LM6-115R	(AF333968) glutamine synthetase [Aspergillus nidulans]	303	8e-82
LM6-195R	MALATE SYNTHASE, GLYOXYSOMAL pir||S17773 malate synthase (EC 4.1.3.2) - Emericella nidulans	303	3e-94
LM9-336	(AL670002) probable autophagy protein AUT7 [Neurospora crassa]	233	2e-60
LM9-257R	Alcohol dehydrogenase I pir||T49440 alcohol dehydrogenase Neurospora crassa emb|CAB91241.1|	231	5e-60
LM6-269R	(AF226997) methionine synthase [Cladosporium fulvum]	175	1e-47
LM6-291R	(AF226997) methionine synthase [Cladosporium fulvum]	142	1e-39
LM6-160F	(AJ458939) aromatic amino acid and leucine permease [Penicillium chrysogenum]	132	3e-30
LM6-136R	(AF306534) glutamine synthetase [Coccidioides immitis]	167	5e-41
LM9-361F	(M35237) argininosuccinate synthetase (ARG1-semicolon- E.C. 6.8.4.5) [Saccharomyces cerevisiae]	91.3	2e-18
LM9-64	putative aspartate aminotransferase [Schizosaccharomyces pombe] pir||T37507	233	1e-62
	Digestion		
LM9-205	(AF092435) trypsin-like protease [Phaeosphaeria nodorum]	316	4e-85
LM6-233	(AF092435) trypsin-like protease [Phaeosphaeria nodorum]	316	3e-85
LM9-66R	(AF074391) serine protease precursor [Fusarium oxysporum f. sp. lycopersici]	141	5e-33
LM9-85R	(AF074391) serine protease precursor [Fusarium oxysporum f. sp. lycopersici]	105	2e-24
LM9-240R	(AF074391) serine protease precursor [Fusarium oxysporum f. sp. lycopersici]	105	5e-23
LM9-349R	(L19059) protease [Aspergillus niger]	122	2e-28

Table 3-1. Continued

	Miscellaneous		
LM9-201R	(AL683874) Possible pathogenesis-related protein precursor [Aspergillus fumigatus]	62.8	2e-09
LM6-126R	TOXD protein emb|CAA63129.1| unique to isolates that make the cyclic peptide HC-toxin [Cochliobolus	146	1e-34
LM9-54	TOXD protein emb|CAA63129.1| unique to isolates that make the cyclic peptide HC-toxin [Cochliobolus	131	1e-49
LM9-263R	(AF327521) TRI15 [Fusarium sporotrichioides]	105	4e-22
LM9-376R	TOXD protein emb|CAA63129.1| unique to isolates that make the cyclic peptide HC-toxin [Cochliobolus	90.1	1e-19

Table 3-2. A list of ESTs of interest from the ILS library, presented here are candidates for defense or pathogenicity genes with descriptions of their closest similarity matches.

Clone name	Description of sequences with closest similarity	Score	e-value
	putative LRR type resistance genes		
ILS-22	"leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana] pir||C84726	536	e-151
ILS-891R	"leucine rich repeat protein (LRP), putative [Arabidopsis thaliana] gb|AAG40341.1|	306	2e-82
ILS-1497R	leucine rich repeat protein family [Arabidopsis thaliana] pir||D84648 probable disease resistance protein	288	4e-77
ILS-1392	"disease resistance protein (CC-NBS-LRR class), putative [Arabidopsis thaliana] pir||T48468 disease resistance-	238	7e-62
ILS-880R	leucine-rich repeat protein [x Citrofortunella mitis]	238	1e-77
ILS-880F	"leucine rich repeat protein (LRP), putative [Arabidopsis thaliana] gb|AAG40341.1|AF324989_1 AT5g21090	75	2e-41
	digestive enzymes/digestive enzyme inhibitors		
ILS-1068	senescence-specific cysteine protease [Brassica napus]	642	0
ILS-796	ATP-dependent Clp protease proteolytic subunit (ClpR1) [Arabidopsis thaliana] pir||T52451	582	e-165
ILS-292	glycosyl hydrolase family 19 (class IV chitinase) [Arabidopsis thaliana] pir||T47601	536	e-151
ILS-473	ATP-dependent Clp protease proteolytic subunit (ClpP6) [Arabidopsis thaliana] pir||C86251 hypothetical protein	455	e-127
ILS-1336	ATP-dependent Clp protease proteolytic subunit (ClpP4) [Arabidopsis thaliana] gb|AAK68772.1|	332	5e-90
ILS-1371R	senescence-specific cysteine protease [Brassica napus]	343	7e-94
ILS-641R	xyloglucan-specific fungal endoglucanase inhibitor protein precursor [Lycopersicon esculentum]	261	5E-69
ILS-58	protein trypsin inhibitor T18K17.7 [imported] - Arabidopsis thaliana gb|AAG52121.1|	277	1e-73
ILS-84R	polygalacturonase inhibitor protein [Brassica napus] gb|AAM94870.2|	259	1e-68
ILS-520R	"probable beta-1,3-glucanase [imported] - Arabidopsis thaliana"	258	2e-68
ILS-685	protein trypsin inhibitor T18K17.7 [imported] - Arabidopsis thaliana gb|AAG52121.1|	230	1e-59
ILS-222R	chitinase class 4-like protein [Brassica napus]	190	4e-48
ILS-898	"beta-1,3-glucanase"	150	2e-35
ILS-1371F	senescence-specific cysteine protease [Brassica napus]	150	5e-36
ILS-1381F	senescence-specific cysteine protease [Brassica napus]	149	1e-35
ILS-596	"ATP-dependent Clp protease proteolytic subunit (ClpR3), putative [Arabidopsis thaliana] gb|AAM97107.1|	144	5e-34
ILS-101	senescence-specific cysteine protease [Brassica napus]	126	8e-29
ILS-945	trypsin inhibitor -related [Arabidopsis thaliana] gb|AAL91212.1| putative trypsin inhibitor	114	6e-25
ILS-897R	"GLUCAN ENDO-1,3-BETA-GLUCOSIDASE PRECURSOR ((1->3)-BETA-GLUCAN ENDOHYDROLASE)	89	1e-17
ILS-429	arabinogalactan-protein (AGP2) [Arabidopsis thaliana] pir||A84613 hypothetical protein	81	2e-14
ILS-831	glycosyl hydrolase family 19 (chitinase) [Arabidopsis thaliana] pir||G84867 probable endochitinase	72	2e-12
ILS-461	trypsin inhibitor -related [Arabidopsis thaliana] gb|AAL91212.1| putative trypsin inhibitor	238	6e-62

Table3-2. Continued

	Oxygen free radical metabolism		
ILS-658	catalase [Brassica juncea]	638	0.0
ILS-1102	"peroxidase, putative [Arabidopsis thaliana] sp|Q9SMU8|	563	e-159
ILS-195R	catalase 1 [Arabidopsis thaliana] sp|Q96528|CAT1_ARATH Catalase 1 gb|AAM97090.1| expressed protein	382	e-105
ILS-1232R	Superoxide dismutase [Cu-Zn] pir||DSRPZC superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - cabbage	311	4e-84
ILS-195F	catalase 1 [Arabidopsis thaliana]	279	2e-74
ILS-936F	"peroxidase, putative [Arabidopsis thaliana] sp|Q9SB81|	256	1e-81
ILS-438	"peroxidase, putative [Arabidopsis thaliana] sp|Q9SMU8|	237	6e-62
ILS-1156F	catalase 1 [Arabidopsis thaliana]	207	2e-71
ILS-808R	peroxidase [Arabidopsis thaliana]	80	9e-15
	Pathogenesis related/defense related/inducible		
ILS-1464R	Avr9/Cf-9 rapidly elicited protein 264 [Nicotiana tabacum]	284	8E-76
ILS-281	defense-related protein [Brassica carinata]	357	e-112
ILS-1299	pathogenesis-related protein 1a homolog precursor - rape gb|AAB01666.1| PR-1a gb|AAB09587.1|	318	4e-86
ILS-1282	pathogenesis-related protein 1a homolog precursor - rape gb|AAB01666.1| PR-1a gb|AAB09587.1|	318	4e-86
ILS-124	Pathogenesis-related protein R major form precursor pir||JH0230 - common tobacco	317	1e-85
ILS-1327	pathogenesis-related protein 1a homolog precursor - rape gb|AAB01666.1| PR-1a gb|AAB09587.1| [Brassica	315	2e-85
ILS-972	defense-related protein [Brassica carinata]	287	2e-76
ILS-1145	avirulence induced gene (AIG2)related protein [Arabidopsis thaliana] dbj|BAA95746.1|	278	5e-74
ILS-1162	pathogenesis-related protein PR-4 type [Sambucus nigra]	223	2e-57
ILS-1014	pathogenesis-related protein PR-4 type [Sambucus nigra]	223	2e-57
ILS-331	pathogen-inducible alpha-dioxygenase [Nicotiana attenuata]	200	1E-50
ILS-1431	harpin-induced protein 1 family (HIN1) [Arabidopsis thaliana] gb|AAF02134.1|	313	2e-84
ILS-1300	harpin-induced protein 1 family (HIN1) [Arabidopsis thaliana] gb|AAF02134.1|	299	3e-80
ILS-247	gb|AAM67338.1| wound-induced protein WI12 [Arabidopsis th	164	2E-39
ILS-382R	harpin-induced protein 1 family (NDR1/HIN1-like protein 3) [Arabidopsis thaliana] dbj|BAB08954.1|	230	1e-59
ILS-191R	defense-related protein [Arabidopsis thaliana]	93	1e-36
ILS-1155R	pathogenesis-related protein 1 [Brassica juncea]	82	2e-15
ILS-1131	pathogen-inducible alpha-dioxygenase [Nicotiana attenuata]	73	9E-13

Table 3-2. Continued

	Jasmonate/ethylene response		
ILS-460	ethylene-insensitive 3 (EIN3) [Arabidopsis thaliana] gb|AAC49749.1| ethylene-insensitive3	411	e-113
ILS-93R	contains similarity to jasmonate inducible protein [Arabidopsis thaliana]	288	3E-77
ILS-678	Ethylene responsive element binding factor 1 (AtERF1) (EREBP-2 protein) gb|AAL25588.1|	196	8e-50
ILS-87F	contains similarity to jasmonate inducible protein [Arabidopsis thaliana]	175	3e-43
	receptor like protein kinases		
ILS-405	CBL-interacting protein kinase 23 [Arabidopsis thaliana] gb|AAK61494.1|	669	0
ILS-1126R	receptor-like protein kinase-like protein [Arabidopsis thaliana]	210	1E-53
ILS-1126F	receptor-like protein kinase-like protein [Arabidopsis thaliana]	199	4E-50
	senescense asssociated		
ILS-98	nodulin MtN3 family protein [Arabidopsis thaliana] emb|CAC05445.1| senescence-associated protein	325	7e-88
ILS-555R	senescence-associated protein family [Arabidopsis thaliana] pir||H85355	285	3E-76
ILS-337F	[Arabidopsis thaliana] dbj|BAB02704.1| senescencs-related protein-semicolon- dihydroorotate	257	1e-69
ILS-1028	senescence-associated protein -related [Arabidopsis thaliana] pir||E96809 protein F28K19.24	225	8e-58
ILS-223	Nicotiana lesion-inducing like [Arabidopsis thaliana]	199	1e-50
ILS-666	senescence-associated protein family [Arabidopsis thaliana] gb|AAG50101.1|	104	3E-22
ILS-1076	putative senescence-associated protein [Pisum sativum]	87	7E-17
ILS-394	putative senescence-associated protein [Pisum sativum]	85	2E-16
ILS-379	putative senescence-associated protein [Pisum sativum]	64	7e-10
	thaumatin family		
ILS-1091	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	370	e-101
ILS-600	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	370	e-101
ILS-548	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	367	e-100
ILS-1500	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	318	7E-86
ILS-482	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	325	3E-94
ILS-1004	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	254	7E-67
ILS-130R	thaumatin family [Arabidopsis thaliana] gb|AAF79420.1|	208	4E-53

Table 3-2. Continued

	assorted other resistance related genes		
ILS-1531	"cellulose synthase, catalytic subunit (Ath-B) [Arabidopsis thaliana] dbj|BAB09693.1|	357	1E-97
ILS-19	"cinnamyl-alcohol dehydrogenase (CAD), putative [Arabidopsis thaliana] sp|O49482|CAD2_ARATH	301	6E-81
ILS-1010R	H. vulgare Mlo protein homolog [imported] - Arabidopsis thaliana	286	2E-76
ILS-890R	putative nodule-specific protein [Oryza sativa (japonica cultivar-group)]	234	4E-61
	putative pathogenicity genes from L. maculans		
ILS-598	"Exopolygalacturonase precursor (ExoPG) (Galacturan 1,4-alpha-galacturonidase)	294	7E-79
ILS-217	rhamnogalacturonan acetyl esterase [Aspergillus niger]	117	1E-25
ILS-325R	transcription factor [Hypocrea jecorina]	112	6E-24
ILS-409R	zinc cluster transcription factor Fcr1p [Candida albicans]	90	3E-17
ILS-251F	beta-glucosidase 5 [Coccidioides posadasii]	84	8E-16
ILS-1220	glycoprotein CIH1 [Colletotrichum lindemuthianum]	75	1E-16
ILS-1006	mycelial surface antigen CSA1 precursor - yeast (Candida albicans) gb|AAC29486.1|	49	0.00006

that resemble the LRR class of disease resistance genes as well as one putative homolog of the barley *Mlo* resistance gene. In total 89 ESTs, or 7.0% of the ILS library is likely involved in defense responses or pathogenicity.

3.4. Discussion

Two major EST libraries were constructed to gain information about the processes of the *L. maculans*/*B. napus* interaction. The first library, constructed from actively growing mycelial tissue of *L. maculans*, has already increased approximately threefold the amount of coding sequence information available for this organism. The second library, constructed from infected leaf tissue is a resource for new gene discovery in the area of host-pathogen interactions. Four different EST sequences with marked similarity to the LRR class of disease resistance proteins have been identified. These ESTs could potentially represent new resistance genes in *B. napus*. These libraries represent only a small sample of the sequences expressed in either the cultured fungal mycelia or the infected leaf tissue, as evidenced by the low levels of redundancy in the sequence libraries.

A combination of approaches was used to attempt to identify the source organism for each EST in the infected leaf ILS library. A total of 5% of the library is very likely derived from *L. maculans*. This estimate was arrived at primarily by classifying ESTs on the basis of their BLAST results (Figure 3-2). This BLAST analysis still leaves 13% of the library unclassified. In the LM library which is known to consist completely of fungal transcripts, only 57% of ESTs could be identified as fungal based on BLAST results. If one assumes that only 57% of fungal transcripts will be identifiable, then it is probable that a further 4-5% of the unidentified transcripts in the ILS library will have originated from the fungus. An analysis of G+C content and hexamer composition of the ESTs was attempted, and while the mean %G+C for each organism is significantly different, the distributions of G+C contents overlap considerably (Figure 3-3). Plotting the %G+C contents for the ILS library reveals a small secondary peak at the upper end of the distribution that is consistent with a small amount of fungal sequences present (Figure 3-4). While it is not possible to use G+C content to determine organism source with any confidence, almost all putative *L. maculans* ESTs in the ILS library are found at the upper end of the G+C distribution, supporting their BLAST based classification. Another approach, which involves calculating the frequency of hexamers in a sequence was also attempted. This approach had been used previously on libraries from *P. sojae* infected soybean and different *M. trunculata* infected libraries (Hraber and Weller, 2001). A comparison of sequences from *B. napus* and *L. maculans* shows that there is a high rate of false positives (17%), that is, plant sequences misidentified as fungal. In contrast, the error rates reported for this analysis performed on the *G. max/P.sojae* interaction are only 6%. As *B. napus* is an amphidiploid species, the high error rate observed here may be due to variation in hexamer composition between *B. napus* ancestral genomes. Alternatively, this high false positive rate may reflect less difference in codon preference between the two organisms studied here than is present between *G. max* and *P. sojae,* an oomycete. The results of this analysis also support the estimated approximately 10% of ILS sequences being from *L. maculans*, as the number of ESTs classified as fungal increases by 8% over the calibration curve from *B. napus*. Dissimilarity values from this analysis can be used to classify the unidentified ESTs as plant or fungal, but each classification has a 17% chance of being misidentified as fungal derived or a 4% chance of being misidentified as plant derived. A comparison of EST classifications based on BLAST results and hexamer frequencies shows that the two methods are in agreement within the predicted error rates, strengthening support for the putative classifications.

The detectable presence of 5-10% of ILS sequences which are fungal in origin is small when compared to infected tissue libraries created from the *Glycine max/Phytopthera sojae* interaction, where 75% of sequences were pathogen derived (Hraber and Weller, 2001). However, this 5-10% range is in keeping with other host-pathogen libraries in *Medicago trunculata,* rice and potato that contained 15% or less pathogen derived sequences (Hraber and Weller, 2001; Soonok et al., 2001; Ronning et al., 2003). The libraries presented here also contain a significant increase in fungal transcripts compared to the only other *B.napus/.maculans* infection library constructed, which only found one fungal derived transcript out of several hundred (Fristensky et al., 1999). The low representation from *L. maculans* likely reflects a low abundance of fungal hyphae in relation to plant leaf tissue even in an advanced compatible reaction and despite the heavy innoculum load used. The inclusion of *L. maculans* RNA in the subtraction reaction may also have selected against *L. maculans* sequences and artificially lowered the amount of *L. maculans* transcript in the library.

Of the 741 ESTs in the LM library, 374 have similarity to sequences in Genbank with at least a hypothetical function. Of these, 30 sequences with a possible role in pathogenesis could be identified (Table 3-1). These sequences have putative functions that include detoxifying plant phytoalexins or reactive oxygen species. Recent studies of the tomato *Cladosporium fulvum* interaction have reported that the tomato apoplast is a nutrient limiting environment and that numerous genes involved in fungal metabolism, particularly methionine synthase and *Gat*1, may be required for biotrophic growth in this environment (Coleman et al., 1997; Solomon et al., 2000; Solomon and Oliver 2002). In light of this, a number of ESTs involved in general metabolism have been categorized as potential pathogenesis genes since some of these may have a role in providing nutrition during the nutrient limiting *L. maculans* biotrophic growth stage. ESTs encoding products involved in degradation of host materials were also included in the list of transcripts with a possible role in pathogenesis. Finally a few miscellaneous ESTs were identified which could have a role in pathogenesis. Three of these are Tox D homologues. Tox D may be involved in toxin production in the Maize pathogen *Cochliobolus carbonum*, although its role in HC toxin biosynthesis has not been fully established (Pedley and Walton, 2001). The TRI15 gene from *Fusarium sporotrichioides* is also involved in mycotoxin production (Peplow et al., 2003). Each of these sequences was collected from *L. maculans* in culture, thus these genes are most likely involved in normal growth and saprophytic metabolism, not pathogenesis. However, the media used contains plant materials and thus some non-specific plant elicitors may have been present and many of these genes may have dual roles that make them necessary for growth on a host, especially in the nutrient poor apoplast, but dispensible in complete media.

The ILS infected leaf library can be used to identify genes which are more likely to be required for pathogenesis. Seven different transcripts were identified in this library which may have a role in infection (Table 3-2). These include two transcripts which degrade pectins and a β-glucosidase, which are likely involved in the breakdown of plant cell walls. Also identified were two putative transcription factors. One transcript is similar to a proline rich glycoprotein from the fungus *Colletotrichium lindemuthianum* which has been proposed to protect this fungus from recognition by its host (bean) during biotrophic growth (Perfect *et al.*, 1998).

The ILS library also sheds light on plant responses to pathogen attack. ESTs with a probable role in plant defenses are listed in Table 3-2. As mentioned above, four independent putative LRR type resistance genes have been sequenced here. As well, other classes of defense genes are represented including many different ESTs encoding proteins involved in blocking fungal digestive enzymes or attacking fungal cell walls (eg. protease inhibitors, chitinases). Other ESTs are potentially involved in oxygen free radical metabolism, a crucial component of plant defense responses. Still other ESTs are possibly involved in ethylene and jasmonate signaling pathways, both of which have been implicated in triggering plant defense responses (Schenk et al., 2000). On the whole, approximately 10% of the library can be associated with defense responses. A previous *B. napus/L.maculans* infection specific library reported 20% of 277 ESTs sequenced to be involved in defense responses; however the level of sequence redundancy in this library seems to be higher, with only 15 different sequences comprising that 20% of the EST library (Fristensky et al. 1999).

The libraries presented here have increased the available amount of sequence information for *L. maculans*. A total of 551 new cDNAs from the LM library have been identified, this is over three times more than the number of currently available sequences from this organism in Genbank. The genes represented in this set have putative functions that range over all aspects of fungal biology from structural (eg. histones, ribosomal proteins), regulatory (transcription factors, translation initiation factors), metabolic (proteases, alcohol dehydrogenases) and on. Some of these ESTs have a conceivable role in ensuring survival in a hostile host environment and aiding pathogenesis. This library provides a basis for studying genes of interest that can shed light on a wide range of fungal molecular biology. Further, 43% of the library is composed of ESTs having no predictable function. This indicates that there is a large gap in our knowledge of fungal biology and an opportunity to study completely novel genes.

The ILS library similarly provides a resource for gene discovery in both *B. napus* and *L. maculans* and can provide some insight into plant defense and fungal attack. It is clear that even susceptible *B. napus* cultivars mount a defense that includes jasmonate and ethylene signaling pathways, oxygen free radical production, and active fungal attack in the form of chitinases and secondary metabolites. From the fungal side, *L. maculans* produces enzymes involved in active degradation of the host cell wall such as pectinases and a proline rich glycoprotein that has been linked to an ability to evade host detection.

Further, each of these libraries provides a resource for further study of the *B. napus/L.maculans* interaction. The cDNAs described here have been printed onto glass slides for use in microarray analysis (presented in chapter four). This will enable the collection of dynamic expression data for each of these sequences in a variety of different experimental circumstances. Also, a collection of T-DNA tagged mutagenized *L. maculans* has also been created (presented in chapter five). The EST data presented here makes it possible to perform PCR based reverse genetic screens on this mutagenized collection based on the sequence of a particular EST of interest. Identification of putative knockouts in ESTs of interest will enable investigations of gene function.

Works cited

Bent, A.F. (1996) Plant Disease Resistance genes: Function Meets Structure. The Plant Cell 8: 1757-1771.

Coleman M., Henricot B., Arnau J., and Oliver R.P. (1997) Starvation-induced genes of the tomato pathogen Cladosporium fulvum are also induced during growth in planta. Molecular Plant Microbe Interactions 10: 1106-1109.

Fristensky, B., Balcerzak, M., He, D. and Zhang, P. (1999) Expressed sequence tags from the defence response of *Brassica napus* to *Leptosphaeria maculans*. Molecular Plant Pathology online: http://www.bspp.org.uk/mppol/1999/0301FRISTENSKY.

Hammond, K.E. and Lewis, B.G. (1987) The establishment of systemic infection in leaves of oilseed rape by *Leptosphaeria maculans*. Plant Pathology 36: 135-147.

Hraber, P.T., and Weller, J.W. (2001) On the species of origin: diagnosing the source of symbiotic transcripts. Genome Biology 2: research0037.1-0037.14.

Idnurm, A. and Howlett, B.J., (2001) Pathogenicity genes of phytopathogenic fungi. Molecular Plant Pathology 2: 241-255.

Kunkel, B.M. and Brooks, D.M. (2002) Crosstalk between signaling pathways in pathogen defense. Current Opinion in Plant Biology 5: 325-331.

Oliver, R.P., Farman, M.L., Jones J.D.G. and Hammond-Kosack, K.E. (1993) Use of fungal transformants expressing B-glucuronidase activity to detect infection and measure hyphal biomass in infected plant tissues. Molecular Plant Microbe Interactions 6: 521-525.

Pandey, A., and Lewitter, F. (1999) Nucleotide sequence databases: a goldmine for biologists. Trends in Biochemical Sciences 24: 276-280.

Pedley K.F. and Walton, J.D. (2003) Regulation of cyclic peptide biosynthesis in a plant pathogenic fungus by a novel transcription factor. Proceedings of the National Academy of Sciences 98: 14174-14179.

Peplow, A.W., Tag A.G., Garifullina, G.F., and Beremand M.N. (2003) Identification of New Genes Positively Regulated by *Tri10* and a Regulatory Network for Trichothecene Mycotoxin Production. Applied and Environmental Microbiology 69: 2731-2736.

Perfect, S.E., O'Connell, R.J., Green E.F., Doering-Saad C., and Green J.R. (1998) Expression cloning of a fungal praline-rich glycoprotein specific to the biotrophic interface formed in the *Colletotrichum*-bean interaction. The Plant Journal 15: 273-279.

Qutob, D., Hraber, P.T., Sobral B.W.S. and Gijzen, M. (2000) Comparitive analysis of expressed sequences in *Phytophthora sojae*. Plant Physiology 123: 243-253.

Rasmussen U., Giese H., Mikkelsen J.D. (1992a) Induction and purification of chitinase in Brassica napus L. spp. oleifera infected with Phoma lingam. Planta 187: 328-334.

Rasmussen U., Bojsen K., Collinge D.B. (1992b) Cloning and characterization of a pathogen-induced chitinase in Brassica napus. Plant Mol Biol. 20: 277-287.

Rathjen, J.P., Moffet, P. (2003) Early signal transduction event in specific plant disease resistance. Current Opinion in Plant Biology 6: 300-306.

Ronning C.M., Stegalkina S.S., Ascenzi R.A., Bougri O., Hart A.L., Utterbach T.R., Vanaken S.E., Riedmuller S.B., White J.A., Cho J., Pertea G.M., Lee Y., Karamycheva S., Sultana R., Tsai J., Quackenbush J., Griffiths H.M., Restrepo S., Smart C.D., Fry W.E., van der Hoeven R., Tanksley S., Zhang P., Jin H., Yamamoto M.L., Baker B.J., and Buell C.R. (2003) Comparative Analyses of Potato Expressed Sequence Tag Libraries. Plant Physiology 131: 419-429

Schenk, P.M., Kazan, K., Wilson, L., Anderson, J.P., Richmond, T., Somerville, S.C., Manners, J.M. (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proceedings of the National Academy of Sciences 97: 11655-11660.

Solomon P.S., Nielsen P.S., Clark A.J., and Oliver R.P. (2000) Methionine synthase, a gene required for methionine synthesis, is expressed in planta by Cladosporium fulvum. Molecular Plant Pathology 1: 315-323.

Solomon P.S., Oliver R.P. (2002) Evidence that gamma-aminobutyric acid is a major nitrogen source during Cladosporium fulvum infection of tomato. Planta 214: 414-420.

Soonok, K., Il-Pyung, A. and Yong-Hwan, L. (2001) Analysis of genes expressed during rice-*Magnaporthe grisea* interactions. Molecular Plant Microbe Interactions 14: 1340-1346.

.White, O., Dunning, T., Sutton, G., Adams, M., Venter J.C., and Fields C. (1993) A quality control algorithm for DNA sequencing projects. Nucleic Acids Research 21: 3829-3832.

4. Microarray analysis of genes from *Leptosphaeria maculans* and *Brassica napus* expressed over the course of infection

4.1. Introduction

Microarray analysis of gene expression is a relatively recent development in molecular genetic studies. These types of analysis offer a potentially powerful tool to gain information about the changes in expression of a large number of genes simultaneously in response to specific treatments. A set of sequences for study are arrayed on a fixed substrate, commonly glass slides. These slides are then used to probe RNA pools from differently treated tissues to identify between pool differences in the abundance of transcripts corresponding to each arrayed sequence. The data produced can be mined to extract sets of genes that follow similar expression patterns implying that they are coordinately regulated and likely part of the same response network. This approach has been used to examine transcriptional changes in response to both abiotic and biotic factors in plants, including responses to light (Rossel et al. 2002), drought stresses (Seki et al. 2001; Ozturk et al. 2002), cold (Seki et al. 2001), insect herbivory (Reymond et al. 2000), as well as pathogens. In the field of pathogen responses specifically, some studies have examined the effects of purified fungal elicitors such as chitin, *N*-acetylchitinooligosaccharides, or fusicoccin on gene expression (Ramonell et al. 2002; Akimoto-Tomiyama et al. 2003; Frick and Schaller 2002). Others have exogenously applied the signaling molecules salicylic acid, ethylene and methyl jasmonate which have been associated with plant defense pathways as well as intact spores of *Alternaria brassicicola* to study the overlaps and divergences in these signaling networks (Naruska et al. 2003; Schenk et al. 2000). One other study involving plant responses to an intact pathogen has been reported. A tomato mutant line overexpressing the resistance gene *Pto* was compared to a susceptible line after each was challenged with *Pseudomonas syringae* pv *tomato* (*AvrPto*) (Mysore et al. 2003). These studies have been successful in identifying groups of genes that are differentially expressed in response to a pathogen related treatment. Some of the genes identified had previously been associated with various defense responses; however these analyses also indicated that other cellular functions such as photosynthesis, carbohydrate metabolism, and translation were transcriptionally affected (Frick and Schaller 2002; Narusaka et al. 2003; Mysore et al. 2003)

The sequencing of ESTs from free living *Leptosphaeria maculans* and from heavily infected *Brassica napus* leaf tissue described in chapter three provided a platform for identifying novel genes from either the plant or fungus that may have some significant role in the interaction between them. *L. maculans* begins its life cycle *in planta* as a biotroph, growing non-invasively within the leaf mesophyll. Even when *L. maculans* infection is blocked visible signs of defense are often apparent, but not until approximately the eight day after infection (Hammond and Lewis 1987; Chen and Howlett 1996). Very little is known about the mechanisms of survival required by fungi that do not form specialized infection structures for growth within plant tissue. Investigations on one particular noninvasive biotroph, *Cladosporium fulvum*, revealed that conditions of carbon and nitrogen starvation induce a number of genes which are also induced *in planta* (Coleman et al. 1997). It was later demonstrated that a number of amino acids are available in limited quantities during most of the infection, although in compatible infections the amounts of specific amino acids increases (Solomon and Oliver 2001). Interestingly, it was shown that *Gat*1, a *C. fulvum* gene required for γ–aminobutyric acid metabolism is expressed at high levels, at the same time that a tomato gene responsible for the synthesis of γ–aminobutyric acid is also highly expressed in only a compatible infection, suggesting *C. fulvum* manipulates host metabolism (Solomon and Oliver 2002). While these results are beginning to illuminate the physiological processes involved in a non-invasive biotrophic infection, they have been restricted largely to investigations of metabolite levels during infection in one plant-pathogen system.

This chapter reports the study of transcriptional gene expression changes in both the host plant *B. napus*, as well as the pathogen *L. maculans* through the initial asymptomatic phase of infection to the development of the first symptoms. Microarrays containing sequenced cDNAs, from both the host *B. napus* and pathogen *L. maculans* were used to interrogate samples from infected resistant leaf tissues at a series of time points covering the biotrophic growth phase and the switch to symptomatic host defense responses. A total of 76 genes, only five of which were fungal in origin were identified as having a significant change in expression level at one of the time points examined. Some of these genes could be grouped into six reproducible clusters based on shared patterns of expression. The possible biological roles of these genes are discussed.

4.2. Methods and materials

4.2.1. Plant cultivation, spore production and inoculation

Individual plants of the blackleg resistant *Brassica napus* doubled haploid line DH12075 were grown in separate 4 inch pots in a growth chamber with a 16hr. light period, 22° C daytime temperature and a 16° C nighttime temperature. *L. maculans* isolate PL86-12 was prepared as described in chapter two. When plants had grown 5-6 leaves, the third youngest leaf was inoculated with spore suspension. Only one leaf per plant was infected. Inoculations were performed as described in chapter three. Infected leaves were harvested at 0.5, 1, 2, 4, and 8 days after infection. Also, leaves from uninfected plants were harvested at the time of infection for use as time point zero controls.

4.2.2. RNA extraction and labeling

Infected leaves were flash frozen in liquid N_2 at harvest and stored at -80°C until RNA extraction. Frozen tissue was ground to a fine powder in a pre-chilled mortar with a pestle and RNA was isolated using TRIzol reagent (Invitrogen Life Technologies). Total RNA was checked for quality on agarose gels. 50μg of total RNA was labeled with either Cy3 or Cy5 conjugated dUTP (Amersham Biosciences) in a first strand cDNA synthesis reaction also containing 10mM dATP, dCTP, dGTP and 2mM dTTP; 100μM oligo $dT_{(18)}VN$; 0.1M DTT; RNAse inhibitor; and Powerscript reverse transcriptase (BD Biosciences Clontech). The labeling reaction proceeded for 2 hrs. at 42° C. EDTA and

NaOH were then added to the reaction and the mixture was heated to 65° C for 10 min.. Labeled targets were purified using microcon YM-30 spin columns (Millipore), before being resuspended in ArrayHyb microarray hybridization buffer (Sigma-Aldrich).

4.2.3. Preparation of cDNA microarray slides

Inserts from EST libraries developed from *in vitro* cultured *L. maculans* and from *B. napus* leaf tissue infected with *L. maculans* (described in chapter three) were amplified from original insert bearing pSPORT1 plasmids with M13 forward and reverse primers. Amplified cDNAs were electrophoresed on agarose gels to confirm quality and the presence of a single band. Amplified cDNAs were purified using MinElute 96 UF PCR purification kits (Qiagen). cDNAs were printed onto ArrayIT superamine coated glass slides (Telechem International, Inc.) using the OmniGrid 100 (Genemachines). Each feature was spotted in triplicate. Also included on the slides were a set of Alien spike control spots (Stratagene), 3XSSC spots, as well as oligo dA, Cot1 DNA and salmon sperm non specific binding controls.

4.2.4. Hybridization

Printed slides were rehydrated by exposing spots to steam briefly and snap drying on a 65° C heat block. Rehydrated slides were immobilized by UV crosslinking (120mJ). Following immobilization, slides were washed once in 1% SDS and once in 100% ethanol. Labeled target resuspended in 12 µl ArrayHyb microarray hybridization buffer (Sigma) was denatured at 95° C for 5 min. and deposited on the microarray slide according to the manufacturer's instructions. The hybridization buffer/target mixture was enclosed by a 2cmx2cm plastic Hybri-slip (Sigma). Hybridization occurred overnight at 65° C in a homemade hybridization chamber consisting of a sealed horizontal 50mL Falcon tube containing a 1.5cmx4cm strip of paper towel soaked in dH_20. After hybridization, slides were washed once in 1X SSC, 0.03% SDS at 60° C, once in 0.2X SSC and once in 0.5X SSC then dried by centrifugation. Each wash lasted five minutes.

4.2.5. Data collection

After drying, slides were scanned with the arrayWoRx^c biochip reader (Applied Precision). For each slide, images of each channel (Cy3, Cy5) were produced in the TIFF file format. These images were imported into TIGR_Spotfinder v.2.2.3 software, part of the TM4 suite of free software developed by The Institute for Genomic Research and available at http://www.tigr.org/software/tm4/. Spots deemed bad due to low signal or irregular spot shape were flagged and excluded from later analysis. Local background corrected spot intensity values were exported from TIGR_Spotfinder. Intensity data were averaged over the three spots for each slide feature and these averaged intensities were normalized using the lowess locally weighted regression function in TIGR MIDAS. Significantly up or downregulated genes were identified using Significance Analysis of Microarrays (SAM) software (http://www-stat.stanford.edu/~tibs/SAM/). Briefly, for each time point comparison both technical replicates of both biological replicates (a total of four replicates per time point comparison) were used in a one class response test to identify genes that had a fold change significantly different from zero. All four replicates for each time point comparison were then averaged. One class response tests in the significance analysis of microarrays (SAM) software application were used to identify genes significantly upregulated or downregulated between adjacent time points. Two class unpaired tests were used to identify genes with a significant expression change between other time points using the intervening time points as a common reference.

Averaged data from the set of genes exhibiting a significant fold change in at least one time point comparison were used to identify clusters of genes with similar expression patterns. Clustering was performed using TIGR Multi Experiment Viewer (MEV) software, also part of the TM4 suite. Expression patterns were grouped using the kendalls tau distance metric. Both k-means clustering and hierarchical clustering methods were employed to come up with general consensus clusters. Expression information along with cluster affiliation and SAM results for each gene was incorporated into the Microsft Access databse LESTer described in chapter three.

4.3. Results

Glass slide cDNA microarrays were constructed to contain all ESTs described in chapter three. Approximately one third of these ESTs were derived from *in vitro* grown *L. maculans* (LM), while the remaining two thirds were derived from an infected *B. napus* leaf (ILS) and represent transcripts from both plant and fungus. A total of 6912 features, representing 2304 triplicated cDNA and control spots were printed. These slides were used to interrogate RNA samples harvested at six different time points (0, 0.5, 1, 2, 4, and 8 days) from blackleg resistant DH12075 leaf tissue infected with *L. maculans* PL86-12 spores. In order to reduce the resources necessary for conducting a time course experiment, a spoke strategy involving a common reference for each time point was rejected in favor of a wheel strategy which compares each time point directly to its adjacent time point (Figure 4-1).

4.3.1. Microarray hybridization replicates

A large number of spots in the t=0, t= 0.5 day comparison and in the t=0.5 day, t= 1 day comparison were flagged as bad and excluded from analysis due to high background levels on these slides, possibly due to insufficient washing. After averaging triplicated spots only 506 genes from the t=0, t=0.5 day comparisons and 594 genes from the t=0.5, t=1 day comparisons were left for which useful data was obtained. The other four comparisons were very consistent, yielding usable data representing between 1113 and 1167 genes. For each time point comparison replicated samples were collected from two independent infected plants. For each biological replicate, extracted RNA was used in two independent dye-flipped hybridizations, for a total of four replicates. Log base 2 expression ratios for each feature exhibited poor correlation between replicates; between some replicates negative correlations were observed (Table 4-1). One case in particular, the comparison between the uninfected control and the half day time point showed a very strong negative correlation between technical replicates. Spot intensity data from these slides were double checked and found to be accurate, although an initial mishandling of RNA samples for one group of technical replicates could not be ruled out. However a simple experimental error would not account for other near zero and weak negative correlations. An example of correlations between replicates is given for the half day and one day timepoint comparisons (Figure 4-2). There was no relation between variation in spot expression ratios and spot signal intensity.

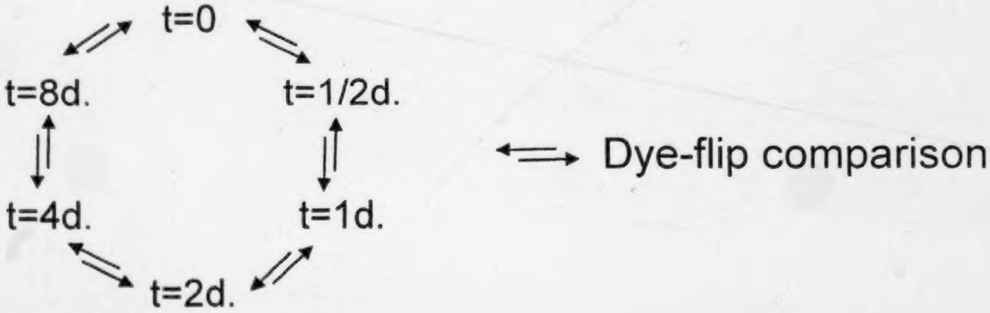


Figure 4-1. Illustration of the experimental strategy for time point comparisons. Six time points after infection spanning the biotrophic growth phase of *L. maculans* to the appearance of symptoms on plants were chosen for study. Each time point was compared directly to the point preceding and following. Four replicates for each comparison were used, two technical replicates in a dye-flip comparison within two biological replicates.

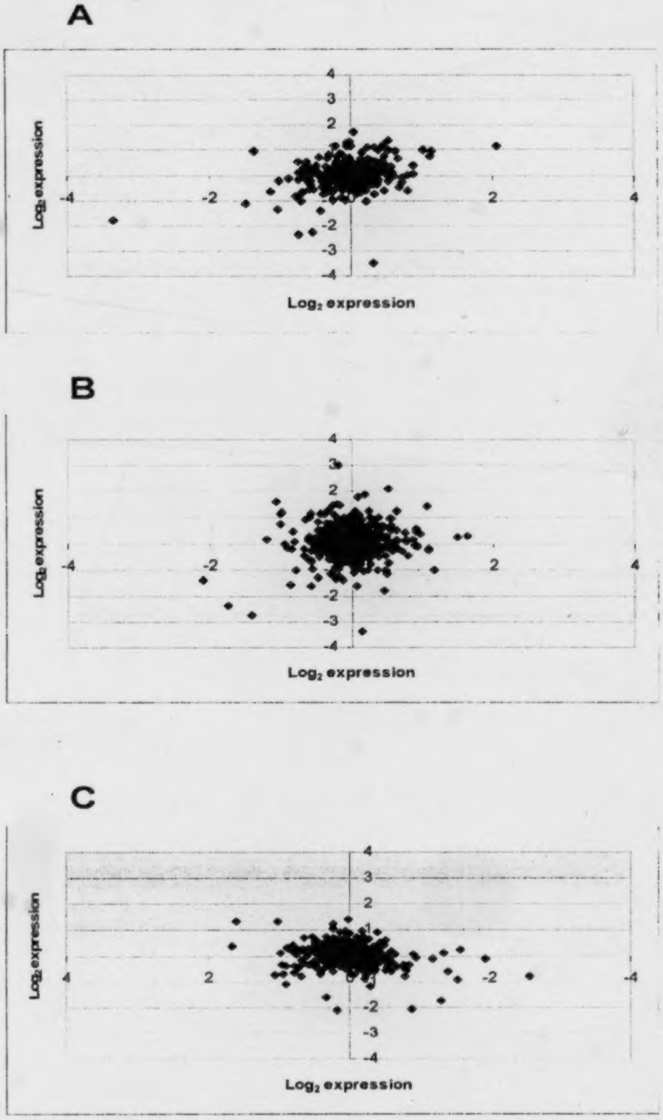


Figure 4-2. Correlation between Log_2 ratios for different replicate slides of the half day and one day timepoint comparisons. The Log_2 expression ratios for each gene one replicate slide are plotted against their corresponding Log_2 expression ratios for another replicate slide. **A** technical flip dye replicate 1, showing the correlation between expression ratios for a dye flip replication involving the same RNA samples. **B** technical dye flip replicate 2 **C** Biological replicate, showing the averaged Log_2 expression ratios of one dye flip technical replicate plotted against the averaged Log_2 expression ratios of a dye flip technical replicate from an independent RNA sample collected under the same biological conditions.

Table 4-1. Pearson correlation coefficients (r) for comparisons of expression ratios between replicates.

Comparison	technical replicate		biological replicate
	1	2	
0 d vs. 0.5 d	0.2019	0.197871	-0.332333
0.5 d vs. 1 d	0.312903	0.106626	0.228927
1 d vs. 2 d	0.43292	0.222725	-0.184712
2 d vs. 4 d	0.180894	-0.092132	0.027055
4 d vs. 8 d	0.295687	0.251585	0.25045
8 d vs. 0 d	0.456855	-0.217463	0.087418

4.3.2. Identification of genes with significant expression changes

A total of 76 genes were found to have significant changes in expression in at least one comparison. A breakdown of the number of genes determined to be significantly regulated at each time point, and the median false discovery rate (a measure of confidence indicating the proportion of identified genes that are deemed significant due to chance) for each analysis are presented in Table 4-2. Genes exhibiting a significant change in expression are presented in Table 4-3 with their probable function based on BLAST search results and mean Log_2 expression ratios for each time point comparison plus or minus one standard deviation.

Despite making up approximately 1/3 of the features arrayed on the microarray slides, only 5 of the genes identified by SAM are fungal in origin. Of these, four (clone names beginning with LM) have no predicted function. The remaining fungal gene, ILS-1227 is similar to a large subunit ribosomal RNA gene.

Proportionally, most of the genes identified as regulated were from the zero and half day comparisons, four day and eight day comparisons or the eight day and zero comparisons, expression patterns did not appear to change drastically between 1 and 4 days. These genes can be grouped into functional categories. Of the 76 clones identified, 9 (11.8%) are similar to genes known to be light regulated or involved in photosynthesis. All of the light regulated genes that could be measured exhibited a decrease in expression at 0.5 days after infection and all but one measured showed an increase in expression again at 1 day after infection. This may reflect a diurnal effect; the 0.5 day sample was

Table 4-2. Number of genes determined to be significantly upregulated or downregulated at the given false discovery rate for each time point comparison according to the Significance Analysis of Microarrays program.

Time Point comparison	Upregulated	Downregulated	Median false discovery rate (%)
0 vs 0.5 d	0	7	12.6
0.5 d vs 1 d	1	9	6.7
1 d vs 2 d	0	10	8.8
2 d vs 4 d	7	1	9.5
4 d vs 8 d	7	21	3.1
8 d vs 0	9	10	5
0.5 d vs 2 d	0	3	31
2 d vs 8 d	3	6	10.2
4 d vs 0	0	9	11.1

harvested approximately one hour after lights in the growth cabinet had been turned off. All other time point samples were harvested at the same time of day as the uninfected t=0 control in full light. All but one of the genes shows a decrease in expression between 4 and 8 days post infection and 6 of the 9 have higher expression levels in the t=0 uninfected control than in the 8 day post infection sample (Table 4-3). At 8 days after infection, symptoms begin to manifest themselves visibly on the leaf surface. At this point infected areas become chlorotic and cell death generally begins to occur. Of the genes showing a significant expression change either increased or decreased, 15 (19.7%) can be categorized as being involved in general metabolism or housekeeping. Seven of these exhibit a decrease in their expression to 8 days, the point at which chlorotic patches form and cell death begins to occur (Table 4-3). This decline likely reflects a reduction in metabolism linked to the initiation of cell death. Two other genes which show similarity to a glucose-6-phosphate/phosphate translocator (AAC28500.1), and a nitrate reductase (D38219.1) increase their expression between 2 and 8 days post infection. It is unclear why these genes should exhibit an increase in expression at this time although they may be involved in the mobilization of nutrients from dying tissue.

General stress response genes have also been identified by SAM, these include sequences similar to a dehydrin like protein (P30185), the cold regulated protein Cor25 (AAO48424.1) and a drought induced protein (X65637.1) all of them show a decrease in expression at 0.5 or 1d after infection.

Defense and senescence related genes make up 26.3% of the genes identified as significantly changed in their expression at some point after infection. These include 6

Table 4-3. Genes exhibiting a significant change in expression. Genes are categorized broadly by probable function. The gene expression cluster to which each EST is assigned is indicated. Expression ratios highlighted are those that were identified by SAM as being significant. The other category refers to expression changes between two nonadjacent time points deemed to be significant.

Clone name	probable function	cluster	Log$_2$ expression ratios for each time point comparison						
			0 vs 0.5	0.5 vs 1	1 vs 2	2 vs 4	4 vs 8	8 vs 0	other
Photosynthesis/ Light dependent									
ILS-1180F	chlorophyll a/b-binding protein	Red			-0.33±0.80	-0.91±0.14	-1.08±0.77	1.03±1.51	2d_vs_8d
ILS-1341	photosystem II D1 protein	Blue		-1.22±0.75	-0.81±1.34	0.43±0.22	-0.47±1.15	-0.25±0.60	
ILS-1401R	chloroplast FtsH protease	Pale Blue	0.61±0.20	0.20±0.34	-0.02±0.17	-0.13±0.14	0.18±0.29	0.10±0.47	
ILS-1517	chlorophyll a/b binding protein	Red			-0.06±0.22	-0.73±0.74	-1.19±0.33	0.80±0.40	2d_vs_8d
ILS-830F	Ribulose bisphosphate carboxylase small	Red	-1.97±0.90	1.51±0.48	0.49±0.67	-0.37±0.51	-0.70±0.06	1.24±0.83	
ILS-883	ferrodoxin, chloroplast	Red			-0.08±0.33	-0.13±0.26	-0.77±0.24	-0.09±0.14	2d_vs_8d
ILS-413	NAD(P)H-quinone oxidoreductase chain 1	Green		0.45±0.37	-0.26±0.27	0.24±0.12	-1.12±1.08	0.85±0.27	
ILS-103F	ATP synthase gamma chain 1, chloroplast	Pale Blue	-0.77±0.35	1.16±0.67	0.16±0.71	-0.78±0.57	-0.17±0.22	0.03±0.57	
ILS-1255	ATP synthase CF1	Red		0.79±0.78	-0.11±0.50	-0.25±0.71	-1.36±0.70	1.50±0.29	
General metabolism/housekeeping									
ILS-1129	histone H1-1	Blue			-0.64±0.05	0.24±0.34	-0.24±0.74	-0.11±0.21	
ILS-1227	Fungal large subunit ribosomal RNA gene	Purple	0.79±3.17	0.05±1.25	-0.32±1.21	-0.07±0.93	-0.51±0.11	-1.70±0.94	4d_vs_0d
ILS-1293R	GTP-binding protein, putative	Orange	-1.06±1.02	0.32±0.36	-0.04±.70	-0.26±0.35	-0.14±0.16	-0.17±0.46	
ILS-1303	glucose-6-phosphate/phosphate translocator	Blue	0.34±0.18		0.08±0.33	0.53±0.15	0.03±0.10	0.46±0.93	
ILS-1318	expansin 10 (EXP10)	Red		-0.01±0.22	1.04±.74	-0.46±0.35	-1.17±0.70	0.48±0.41	2d_vs_8d
ILS-1350	NADH dehydrogenase subunit J	Yellow		0.38±0.55	-0.01±0.48		-0.92±0.54	1.36±0.41	
ILS-1429	putative auxin-induced protein	Purple			-0.23±0.31	-0.05±0.5	0.17±0.54	-0.76±0.17	
ILS-401	acyl-CoA synthetase	Purple	-0.59±0.32	-0.35±0.08	-0.38±0.51	0.26±0.43	0.65±0.40	0.03±0.33	
ILS-350F	dTDP-glucose 4-6-dehydratase h	Red			0.20±0.76	-0.99±1.16	-1.83±0.67	0.64±1.03	2d_vs_8d
ILS-478R	endo-polygalacturonase	Purple		-0.38±0.05	0.00±0.14	0.18±0.19	0.30±0.19	-0.32±0.25	
ILS-1211	calmodulin	Green			-0.31±0.22	1.03±0.75	-0.58±0.13	1.10±0.34	
ILS-797F	nitrate reductase	Purple			0.04±0.34	0.40±0.39	0.79±0.39	-0.14±0.43	2d_vs_8d
ILS-840F	DNA-directed RNA polymerase alpha chain	Green	-0.23±0.21	0.41±0.08	-0.19±0.23	-0.18±0.15	-0.61±0.41	0.87±0.14	
ILS-1330	putative RING zinc finger protein	No				-0.19±0.40	-0.74±0.24		
ILS-970	NAM (no apical meristem)-like protein	Blue		0.31±0.02	-0.50±0.71	0.52±0.35	-0.62±0.22	0.31±0.58	
senescence related									
ILS-1186	metallothionein-like protein (MT1)	Purple	0.56±0.95	-0.62±0.06	-0.42±0.19	0.09±0.30	0.84±0.38	-1.05±0.53	0.5d_vs_2d
ILS-1526	metallothionein-like protein (MT1)	Purple				0.95±1.35	2.49±1.21	-1.60±1.02	2d_vs_8d
ILS-321	Metallothionein-like protein	No						-1.81±1.45	
ILS-692	metallothionein-like protein (MT1)	Purple		-1.51±0.41	-0.31±0.39	0.49±0.45	1.19±0.32	-1.49±0.79	0.5d_vs_2d
ILS-337	dehydrogenase -related; senescence-related	No			-1.10±0.73				
ILS-98	a senescence-associated protein (SAG29)	Turquoise	-0.65±0.08		0.03±0.20	-0.05±0.54	0.34±0.63		

Table 4-3 continued

defense related									
ILS-1232R	Superoxide dismutase [Cu-Zn]	No Cluster		-0.14±0.03	-0.35±0.76	0.23±0.21	-0.18±0.16	-0.28±0.07	4d_vs_0d
ILS-124	thaumatin-like protein PR-5a	Yellow				-0.49±0.71	-1.37±0.97	0.31±0.62	
ILS-1282	pathogenesis-related protein PR1	Red			0.06±0.96	-0.38±1.07	-2.35±0.50	0.27±0.71	2d_vs_8d
ILS-1327	pathogenesis-related protein PR1	Red		0.31±0.90	0.56±1.12	-0.19±1.28	-2.02±0.73	-0.10±0.81	2d_vs_8d
ILS-1531	cellulose synthase, catalytic subunit	Pale Blue			-0.71±0.31	-0.85±1.11	0.12±0.98	0.36±0.48	
ILS-191R	defense-related protein	Purple			-1.02±0.52	0.43±0.66	1.02±1.17		
ILS-466F	beta-1,3-glucanase bg3	Green		0.57±0.47	-0.08±1.31	-0.36±1.00	-0.79±0.39	-0.21±0.51	
ILS-866F	lipase SIL1	Blue			0.14±0.30	0.53±0.15	-0.21±0.35	0.05±0.17	
ILS-873	glutathione transferase, putative	Purple				0.28±0.34	0.29±0.43	-0.97±0.25	
ILS-936F	peroxidase, putative	No Cluster		0.09±0.59	-0.46±0.20	-0.33±0.44	0.55±1.07	0.83±0.21	
ILS-958	glucan endo-1,3-beta-D-glucosidase	No Cluster			-0.62±0.58		-1.36±0.79		
ILS-548	thaumatin family	Red			0.00±0.58	-0.83±1.09	-0.65±0.31		
ILS-658	catalase	Turquoise		-0.64±1.82	-0.45±0.45	-0.58±0.71	0.71±0.16	-0.09±0.41	
ILS-897F	glucan endo-1,3-beta-d-glucosidase	Green			0.17±1.36	-0.16±0.80	-0.94±0.47	0.71±1.04	
General Stress reponse									
ILS-734R	dehydrin RAB18-like protein	Orange	0.21±0.43	-0.70±0.17	0.05±0.37	-0.09±0.24	-0.02±0.16	-0.20±0.04	
ILS-789	cold-regulated protein Cor25	Blue		-1.48±0.76	-0.59±1.23	0.69±0.75	0.30±0.20	0.42±0.77	
ILS-990R	BnD22 drought-induced protein	Purple		-0.24±0.29	-0.60±0.06	0.13±0.42	-0.05±0.61	-0.30±0.28	0.5d_vs_2d

Table 4-3 continued

Unknown									
ILS-1071	unknown	Blue	-0.23±0.34		-0.19±0.23	0.54±0.71	-0.05±0.56	0.27±0.31	
ILS-1075	unknown	Orange		0.14±1.15	0.41±1.17	-0.53±0.58	-0.54±0.46	-1.42±0.92	4d_vs_0d
ILS-1095F	hypothetical protein	Gold			-1.07±0.81		-0.16±0.19	0.24±0.73	
ILS-1501	unknown	Red	-0.31±0.38	0.53±0.43	0.20±0.25	-0.13±0.13	-0.61±0.18	0.37±0.18	2d_vs_8d
ILS-1514	unknown	Purple				0.46±0.43	1.25±0.58	-0.76±0.95	
ILS-1516	expressed protein	Red			0.03±0.14	-0.25±0.86	-0.21±0.15	0.64±0.10	
ILS-1528	unknown	Purple			-0.32±0.21	-0.01±0.52	0.97±0.57	-0.83±0.38	
ILS-188F	unknown	Orange		0.17±0.77	0.57±0.67	-0.68±0.92	-0.95±0.71	-1.15±0.86	4d_vs_0d
ILS-227	unknown	Turquoise			-0.39±0.96	-0.73±1.36	1.06±0.50		
ILS-365	unknown	No Cluster					1.05±0.49		
ILS-373	unknown	No Cluster	-0.49±1.04	-0.42±0.10					
ILS-507	Hypothetical protein	Purple	0.10±0.09		0.00±0.51	0.49±0.15	0.23±0.89	-0.02±0.36	
ILS-525	unknown	Orange		0.00±0.79	0.38±0.49	-0.12±0.46	-1.17±0.68	-1.59±0.74	4d_vs_0d
ILS-537	unknown	Turquoise			-0.25±0.37		0.84±0.19	0.27±0.71	
ILS-665	unknown	No Cluster			-0.60±0.22				
ILS-676	unknown	Blue		-0.55±0.19	-0.31±0.47	0.24±0.40	-0.49±1.07	-0.02±0.68	
ILS-749	unknown	Purple		-0.48±0.27	0.06±0.50	0.30±0.15	0.18±0.42	-0.57±0.08	
ILS-787	unknown	Green			-0.35±0.58	-0.17±0.42	-1.00±0.91	1.00±0.33	
ILS-833	unknown	Red			1.10±1.03	-0.26±0.16		0.71±0.12	
ILS-834	unknown	Gold			-0.67±0.31	0.06±0.56	-0.41±1.05		
ILS-905	unknown	No Cluster					-0.50±0.13	0.30±0.3	
ILS-930F	unknown	Gold		-0.36±0.26	-0.75±0.19	0.11±0.39	-0.44±1.64		
ILS-959	unknown	Red			-0.13±0.65	-0.61±1.18	-1.25±0.51	0.64±0.47	
ILS-963	unknown	Orange			0.23±.51	0.02±0.4	-0.43±0.42	-0.60±0.41	4d_vs_0d
LM6-306	unknown	Purple			-0.16±0.16	0.49±0.12	0.02±0.26	-0.36±0.41	
LM6-329F	unknown	Blue			-0.59±0.13	0.19±0.06	-0.09±0.10	0.08±0.41	
LM6-371	unknown	No Cluster	-0.57±0.21	-0.07±0.60	-0.19±0.69	0.03±0.40	0.18±0.41	0.45±0.61	
LM9-298	unknown	Orange				0.68±0.27	-0.01±0.14	-0.34±0.45	

Expression ratios highlighted have been identified as being significantly different from a Log_2 ratio of 0, representing a null hypothesis of no change in expression between the two time points compared.

genes putatively related to senescence, 4 sequences similar to metallothionein like proteins (AF458412.1; AAC60575.1), and 2 generic senescence related sequences (BAB02704.1; and AAK76623.1). These appear to be down regulated initially after infection with an increase in expression to 8 days post infection. Sequences similar to genes with a role in oxygen free radical metabolism comprise three of these defense related genes (DSRPZC; Q9SB81; AAD17935.1). Other defense related genes include those showing similarity to beta-glucosidases and a beta-glucanase (AAA32755.1; CAA03908.1; and T45802) potentially involved in enzymatic digestion of invading fungal tissue, although these sequences exhibit a decrease in expression over the course of an infection.

A large proportion (36.8%) of the genes significantly up or down regulated in response to infection have no known function or no similarity at all to previously studied genes.

4.3.3. Gene clustering based on expression patterns

Averaged expression data for all genes that were identified as having significant expression changes were imported into TIGR MeV to search for clusters of genes with similar expression patterns. Because a large proportion of data points from the t=0, t= 0.5 day and t=0.5 day, t= 1 day comparisons were missing, these were omitted from the clustering analysis. Both the k-medians clustering support and hierarchical clustering functions were used. The k-medians clustering support function runs the k-medians clustering function multiple times, generating consensus clusters at user specified thresholds from these multiple runs. The 76 genes examined here formed 9 distinct

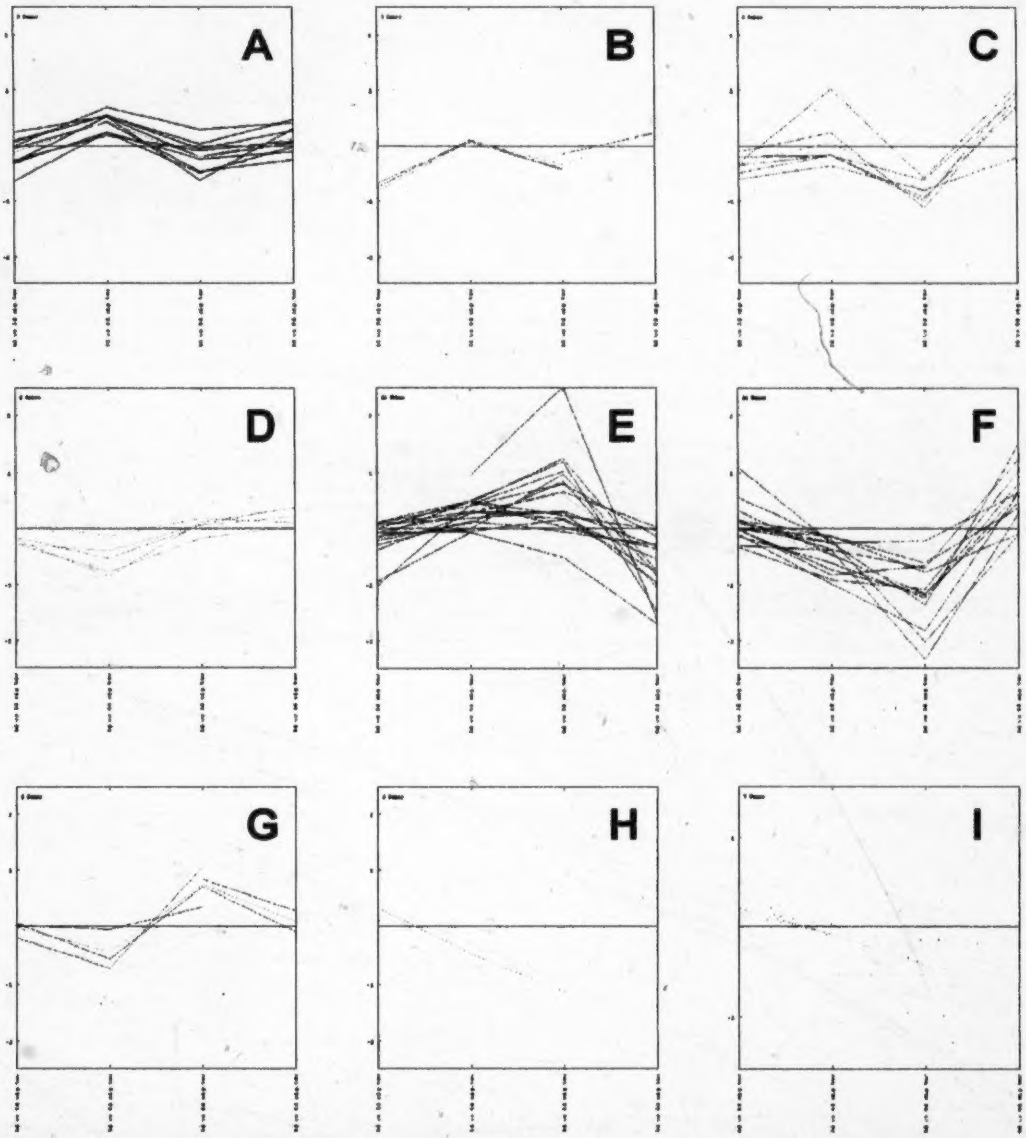


Figure 4-3. Gene expression consensus clusters identified by multiple K-medians clustering runs. All genes identified by SAM as having significant changes in gene expression at one time point after infection were clustered using the KMS function in TIGR MeV with 10 K-medians runs and a co-occurrence threshold of 66% to identify genes showing similar patterns of expression. Each point on the graphs represents the change in gene expression expressed as a Log$_2$ ratio between one time point and the time point preceding it.

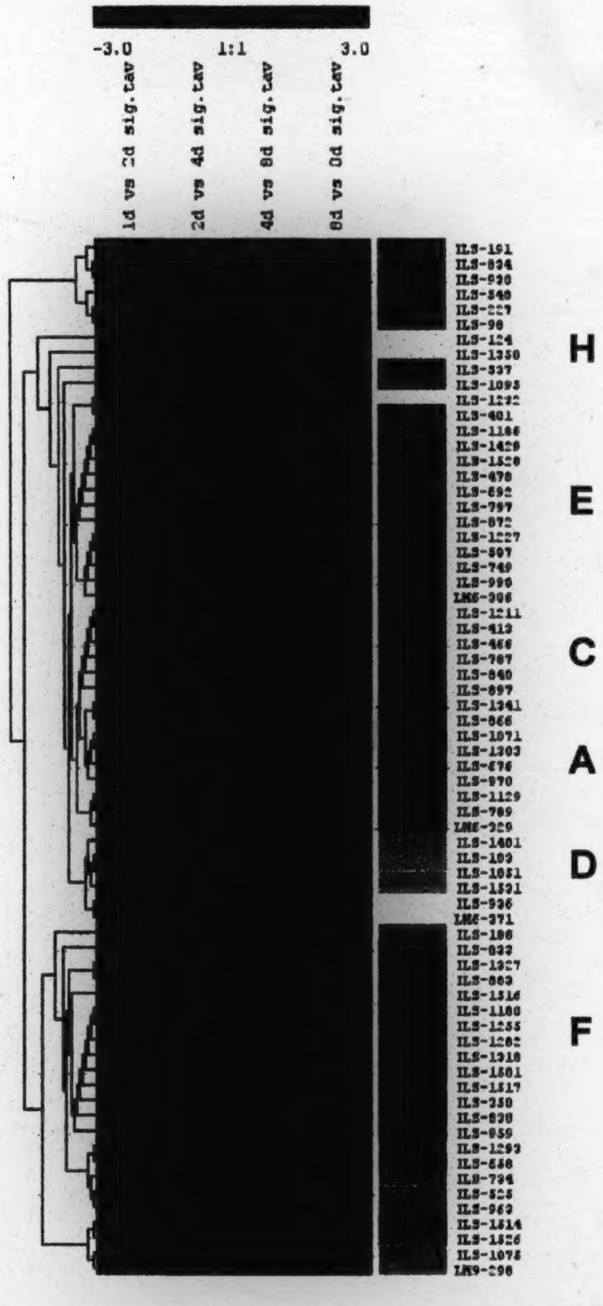


Figure 4-4. Gene expression clusters as identified by hierarchical clustering. All genes identified by SAM as having significant changes in gene expression at one time point after infection were clustered using the HCL function in TIGR MeV using Kendall's Tau distance and average linkage to group genes showing similar patterns of expression. Individual hybridization experiments are identified at the top of the gene expression columns

consensus clusters after 10 k-medians runs with a co-occurrence threshold of 66% (Figure 4-3). Using the hierarchical clustering function, 4 of the previously identified clusters were exactly reproduced, a further two clusters had one gene reassigned and in the remaining 3 clusters two or more genes had been reassigned (Figure 4-4). Overall, the two clustering techniques were consistent.

4.4. Discussion

4.4.1. Consistency of microarray results

One of the primary concerns about using microarray technologies effectively is the reliability of the data produced. A large number of factors can contribute to variability in microarray data. These include differences between gene expression in biological replicates, inconsistencies in RNA quantification, degredation in RNA quality, biases in dye incorporation, spatial variation over slide surfaces, and small differences in hybridization or washing conditions. In this study, a lack of correlation between Log_2 ratios in both technical and biological replicates was observed (Table 4-1). This variation between replicate hybridizations is common in microarray experiments and necessitates careful interpretation of the resulting data (Finklestein et al. 2002). Recently, statistical methods designed specifically for microarray experiments have been developed. One such method is significance analysis of microarrays, which examines gene expression ratios relative to the standard deviation of repeated measurements and establishes a false discovery rate based on permutations of the gene expression data (Tusher et al. 2001). Although no corroborating experiments were performed to confirm the expression changes observed in these microarray hybridizations, a number of the genes identified here have support for there induction or repression in previously published literature.

Out of 2304 features printed on the microarray slides, only 76 (3.3%) were identified as being significantly changed in their expression (Table 4-3). This is not inconsistent with other microarray analyses of pathogen induced gene expression. From a microarray containing 8987 rice genes, 2.9% were identified with altered expression in response to an N-acetylchitooligosachharide elicitor (Akimoto-Tomiyama et al. 2003). A study of chitin elicitation of gene expression in *A. thaliana* identified 3.0% of genes on a microarray containing 2375 features with expression changes grater than three fold (Ramonell et al. 2002). With the same set of 2375 *A. thaliana* genes 8.8% showed a fold change greater than 2.5 when challenged with *Alternaria brassicicola* (Schenk et al. 2000).

4.4.2. Expression of *L. maculans* genes over the course of infection

One of the goals of these experiments was to investigate *L. maculans* gene expression *in planta*. One-third of the cDNAs used to construct these microarrays were collected from cultured *L. maculans*, while two-thirds of the cDNAs used were initially harvested from infected susceptible leaf tissue. A heavily infected susceptible leaf was chosen for library development to maximize the amount of fungal tissue and thus the proportion of fungal transcripts represented. The proportion of fungal genes in this infected leaf library was estimated to only be between 5 and 10% (chapter three). Although fungal transcripts did not appear to be very abundant in susceptible leaf tissue, it was hoped that the microarray technique would be sensitive enough to monitor the expression of *L. maculans* genes during pathogen growth as well as *B. napus* genes activated as part of the host defense response even in resistant leaf tissue. Despite comprising over one third of the features on the microarray slide and representing approximately one third of genes giving a usable signal in each comparison, very few (five of 76) genes from *L. maculans* were identified as being significantly changed in their expression. More disturbingly, the comparisons involving the uninfected sample showed strong signals from both RNA samples for most fungal spots. Fungal genes by definition should not be observed to be expressed in uninfected leaf tissue. None of the non-specific hybridization controls; 3x SSC, Cot1 DNA, polyA DNA or Salmon Sperm DNA exhibited any signal above background. Of the fungal sequences arrayed on these slides, only 1% have any significant similarity to plant sequences (chapter three), making cross hybridization between related genes from the two organisms unlikely. It is still possible that the post-hybridization washes were not sufficient for these slides. The zero and half day comparisons were noted to yield fewer spots with detectable signal as a result of high background signal likely due to insufficient washing. It may also be that the RNA samples were contaminated. Potential contaminants that could hybridize with fungal sequences would be either sequences with similarity to fungal genes or simply generic vector sequences, as the cDNAs generated for deposition on the glass slides were amplified using vector primers and contain short stretches of (108 and 120 base pairs). It was not possible to determine the source of this unexpected hybridization without further testing. The relevance this observation has to the data collected from the *B. napus* genes arrayed on these slides also was not be determined. Performing hybridizations using

RNA derived strictly from *L. maculans* would determine whether or not fungal sequences would also hybridize to plant derived cDNAs.

4.4.3. Expression of *B. napus* genes over the course of infection

The other major goal of this research was to examine *B. napus* gene expression changes over the course of an infection and resistance response. Genes identified here can be broken down into broad functional categories based on their sequence similarity to previously studied genes. Photosynthesis related genes made up nine of the 76 genes identified. One of these genes contains sequence identical to the *B. napus* and *A. thaliana* *psbA* gene encoding the photosystem II D1 protein. Oddly enough, this sequence appears to be down regulated at 1 day as compared to the 0.5 day sample. This is odd because the 0.5 day sample was harvested in the dark, while the 1 day sample was harvested in full light. Previous studies of this photosystem II gene in tobacco demonstrated that it was transcriptionally activated by light (Chun et al. 2001). Closer examination of this sequence revealed that it is chimeric, with the 5' end identical to the *psbA* gene and the 3' end identical to the fungal *Phoma* species 28s large subunit ribosomal RNA gene. Given that the sequence appears to be a chimera it is impossible to determine what the observed decrease in gene expression actually represents. Three other genes, encoding a putative chloroplast FtsH protease, putative ribulose bisphosphate carboxylase small subunit and putative chloroplast located ATP synthase show a decrease in expression at 0.5 days post infection with an increase again at 1 day. This is very likely due to light regulation, as all three genes are known to be transcriptionally stimulated by light (Lindahl et al. 1996; Dedonder et al. 1993; Inohara et al. 1991). The majority of these photosynthesis related genes exhibit decreased expression at 8 days (either a significant decrease between four days and eight, or a significant increase between eight days and the t=0 control). As tissue death and chlorosis occurs around eight days, this is not surprising; decreases in photosynthesis related gene expression have been linked to senescence (Miller et al. 1999).

Genes with a role in general cellular functions and metabolism account for 21.1% of the identified sequences. Given that the samples studied were collected over an eight day period, some of the expression changes noted may reflect processes related to leaf growth and development, and not specifically pathogen response. One example may be the decrease in expression of an endo-polygalacturonase, as this gene is required for the loosening of cell walls to permit growth. Two genes exhibit an increase in expression between two days and four days after infection; these are similar to a glucose-6-phosphate/phosphate translocator and a calmodulin respectively. These also may simply be playing a role in general growth and metabolism.

Two genes appear to be downregulated immediately after infection, these are; a predicted GTP-binding protein and an acyl-CoA synthetase. However the significance of this change in expression is unclear. Of the putative general metabolism/housekeeping genes identified as having a significant expression change, seven showed a decrease in expression at eight days post infection, indicating that they too may reduce their expression as a response to senescence.

A little more than one quarter (26.3%) of the genes identified by this analysis categorized as being involved in pathogen response or leaf senescence based on evidence for their function present ed in the literature. This proportion is slightly higher than that observed in other microarray studies of gene expression changes after pathogen treatment. Typically, 10% of genes identified by these studies are involved in known pathogen defense responses (Schenk et al. 2000; Akimoto-Tomiyama et al. 2003; Narusaka et al. 2003). The larger proportion of defense related genes reported here likely reflects the nature of the arrayed cDNAs, two-thirds of which were collected from an infected leaf library enriched for infection specific transcripts.

Six genes were found that are associated with senescence and follow a similar expression pattern. Four of these are sequences similar to metallothionein proteins. These proteins are known to be involved in senescence responses, both developmental as well as pathogen induced cell death (Butt et al. 1998). Three of the genes categorized as defense response genes had a possible role in oxygen free radical metabolism. The [Cu-Zn] superoxide dismutase is known to be inducible by both virulent and avirulent pathogens in *Arabidopsis* and may function to limit the spread of cell death by reducing superoxide anion levels (Kliebenstein et al. 1999). Catalase is also induced in *Brassica* tissues undergoing senescence (Buchanan-Wollaston and Ainsworth 1997). Surprisingly, a sequence identical to an *Arabidopsis* peroxidase is expressed at higher levels in the t=0 control sample than at eight days after infection. Peroxidases can act as a source of active oxygen species and the downregulation of an *Arabidopsis* peroxidase has been demonstrated to render plants more susceptible to infection by *Pseudomonas syringae* (Bolwell et al. 2002), therefore it might have been expected that the expression of this gene would increase upon infection.

Other genes that have been linked to pathogen response and would be expected to exhibit increased expression after infection actually seem to be downregulated. Seven genes including sequences similar to the pathogenesis related PR-1 and PR-5 proteins as well as beta-1,3-glucanase and 1,3-beta-glucosidase, show a significant drop in expression between four and eight days post infection. Previously, PR-1 had been noted to be strongly induced in *B.* napus two days post infection with an avirulent isolate (Fristensky et al 1999). It is possible that expression increased greatly at one of the initial timepoints for which data is lacking, and the reduction between four and eight days represents the resumption of normal levels of expression, or a reduction in expression due to cell death. A sequence similar to cellulose synthase was observed here to decrease expression significantly between one and two days. It has recently been reported that a reduction in cellulose synthesis may be related to defense against pathogens by triggering jasmonic acid and ethylene synthesis pathways as well as stimulating lignin production (Cano-Delgado et al. 2003). Previously a salicylic acid inducible lipase was shown to also be induced by *P. syringae* infection of *B.rapa* (Lee and Cho 2003). A sequence identical to this gene was also seen here to increase its expression at four days post infection in this study.

Finally a putative glutathione transferase was demonstrated to be induced relative to the uninfected t=0 time point at eight days post infection. Glutathione transferases are a large family of genes, but some members in *A. thaliana* have been shown to exhibit increased expression in response to infection by *Peronospora parasitica* (Wagner et al 2002).

The majority of the genes identified by this microarray analysis (36.8%) have no predictable function, they either have very weak similarities to any sequences in Genbank, or are similar to uncharacterized ‘unknown’ or ‘hypothetical’ proteins. This illustrates one of the advantages of a microarray platform which can identify novel sequences with a possible role in pathogen response that may not have been easily identified by other means. All of the sequences identified here with no similarity to known genes are prime candidates for further analysis to determine their function and relation to pathogen defense pathways.

4.4.4. Identifying genes with similar expression patterns

Two different methods were used to partition significantly induced genes into groups sharing similar patterns of expression changes. Six clusters of genes were consistently identified by both the k-medians and hierarchical clustering methods. The six consistent clusters contain genes from a number of the different functional categories, there does not seem to be a clear correlation between common expression pattern and probable function. Although some genes such as the pathogen related PR genes PR-1 and thaumatin share a common expression pattern, they also cluster together with photosynthesis related genes.

This grouping together of genes involved in different cellular functions based on shared patterns of expression means that it would be unadvisable to attempt to use shared expression patterns to assign putative functions to 'unknown' genes through guilt by association. Clustering methods have been used previously to identify groups of genes that share common regulatory elements (Ramonell et al. 2002). It is probable that by cloning the genomic copies of the ESTs in a cluster and analyzing their upstream sequences, regulatory sequences common to some of these genes could be found.

4.5. Conclusions

The results presented here provide a basis for further study of plant-pathogen interactions. Variation between replicates in these microarray experiments is high, limiting their usefulness as a tool to collect reliable expression data for each gene represented on the array. However, a subset of genes with definite changes in expression that were consistent between replicates could be identified. Although no independent experiments were conducted to confirm the changes in expression levels observed with these microarrays, a number of the sequences identified as being significantly induced/repressed are homologues of genes which are known to be regulated by light, pathogen or senescence pathways. This consistency between the observations presented here and previously published results suggests that these expression changes are real effects. While failing to detect transcriptional changes from *L. maculans* during infection, plant defense response were observed. While some of the changes in transcript level observed here have been documented previously, other observations, such as the reduction in pathogenesis related protein (PR) expression were unexpected and warrant further study. A set of genes showing differential expression were identified which have no currently known function. These should be examined in more detail to discover their function and assess their role in plant defense against pathogens. Finally, sets of genes following similar expression patterns were identified, these coordinately regulated genes can be used to identify novel upstream promoter elements.

Works Cited

Akimoto-Tomiyama C., Sakata K., Yazaki J., Nakamura K., Fujii F., Shimbo K., Yamamoto K., Sasaki T., Kishimoto N., Kikuchi S., and Shibuya N, Minami E. (2003) Rice gene expression in response to N-acetylchitooligosaccharide elicitor: comprehensive analysis by DNA microarray with randomly selected ESTs. Plant Mol Biol. 52: 537-551.

Bolwell G.P., Bindschedler L.V., Blee K.A., Butt V.S., Davies D.R., Gardner S.L., Gerrish C., and Minibayeva F. (2002) The apoplastic oxidative burst in response to biotic stress in plants: a three-component system. J. Exp. Bot. 53: 1367-1376.

Buchanan-Wollaston V. and Ainsworth C. (1997) Leaf senescence in *Brassica napus*: cloning of senescence related genes by subtractive hybridization. Plant Mol Biol. 33: 821-834.

Butt A., Mousley C., Morris K., Beynon J., Can C., Holub E., Greenberg J.T., and Buchanan-Wollaston, S. (1998) Differential expression of a senescence-enhanced metallothionein gene in *Arabidopsis* in response to isolates of *Peronospora parasitica* and *Pseudomonas syringae*. Plant J. 16: 209-221.

Cano-Delgado A., Penfield S., Smith C., Catley M., and Bevan M. (2003) Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. Plant J. 34:351-362.

Chen C.Y. and Howlett B.J. (1996) Rapid necrosis of guard cells is associated with the arrest of fungal growth in leaves of Indian mustard (Brassica juncea) inoculated with avirulent isolates of Leptosphaeria maculans. Physiol Mol Plant Pathol 48: 73-81.

Chun L., Kawakami A., and Christopher D. A. (2001) Phytochrome A mediates blue light and UV-A-dependent chloroplast gene transcription in green leaves. Plant Physiology 125: 1957-1966.

Clark A.J., Bissett K.J., and Oliver R.P. (2003) Investigating the role of polyols in Cladosporium fulvum during growth under hyper-osmotic stress and in planta. Planta 216: 614-619.

Coleman, M., Henricot, B., Arnau, J., and Oliver, R.P. (1997) Starvation-induced genes of the tomato pathogen *Cladosporium fulvum* are also induced during growth in planta. Mol Plant Microbe Interact. 10:1106-1109.

Dedonder A., Rethy R., Fredericq H., Van Montagu M., and Krebbers E. (1993) Arabidopsis rbcS genes are differentially regulated by light. Plant Physiology 101: 801-808.

Finklestein D., Ewing R., Gollub J., Sterky F., Cherry J. M., and Somerville, S. (2002) Microarray data quality analysis: lessons from the AFGC project. Plant Molecular Biology 48: 119-131.

Frick U.B., Schaller A. (2002) cDNA microarray analysis of fusicoccin-induced changes in gene expression in tomato plants. Planta 216: 83-94.

Fristensky, B., Balcerzak, M., He, D. and Zhang, P.(1999) Expressed sequence tags from the defence response of *Brassica napus* to *Leptosphaeria maculans*. Molecular Plant Pathology online: http://www.bspp.org.uk/mppol/1999/0301FRISTENSKY.

Hammond K.E., and Lewis B.G. (1987) The establishment of systemic infection in leaves of oilseed rape by *Leptosphaeria maculans*. Plant pathol. 36: 135-147.

Inohara N., Iwamoto A., Moriyama Y., Shimomura S., Maeda M., Futai M. (1991) Two genes, atpC1 and atpC2, for the gamma subunit of Arabidopsis thaliana chloroplast ATP synthase. Journal of Biological Chemistry 266: 7333-7338.

Kliebenstein D.J., Dietrich R.A., Martin A.C., Last R.L., and Dangl J.L. (1999) LSD1 Regulates Salicylic Acid Induction of CopperZinc Superoxide Dismutase in *Arabidopsis thaliana*. Mol Plant Microbe Interact. 12: 1022-1026.

Lee K.A., and Cho T.J. (2003) Characterization of a salicylic acid- and pathogen-induced lipase-like gene in Chinese cabbage. J Biochem Mol Biol. 36:433-41.

Lindahl M., Tabak S., Cseke L., Pichersky E., Andersson B., and Adam Z. (1996) Identification, characterization, and molecular cloning of a homologue of the bacterial FtsH protease in chloroplasts of higher plants. Journal of Biological Chemistry 271: 29329-29334.

Miller J.D., Arteca R.N., and Pell E.J. (1999) Senescence-Associated Gene Expression during

Ozone-Induced Leaf Senescence in Arabidopsis. Plant Physiol. 120: 1015-1023

Mysore K.S., D'Ascenzo M.D., He X., Martin G.B. (2003) Overexpression of the disease resistance gene Pto in tomato induces gene expression changes similar to immune responses in human and fruitfly. Plant Physiol. 132: 1901-1912.

Narusaka Y, Narusaka M, Seki M, Ishida J, Nakashima M, Kamiya A, Enju A, Sakurai T, Satoh M, Kobayashi M, Tosa Y, Park P, Shinozaki K. (2003) The cDNA microarray analysis using an Arabidopsis pad3 mutant reveals the expression profiles and classification of genes induced by Alternaria brassicicola attack. Plant Cell Physiol. 44: 377-387.

Ozturk Z.N., Talame V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R., Bohnert H.J. (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol. 48: 551-573.

Ramonell K.M., Zhang B., Ewing R.M., Chen Y., Xu D., Stacey G., AND Somerville S. (2002) Microarray analysis of chitin elicitation in *Arabidopsis thaliana*. Mol Plant Path. 3: 301-311.

Reymond P., Weber H., Damond M., Farmer E.E. (2000) Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. Plant Cell 12: 707-720.

Rossel J.B., Wilson I.W., Pogson B.J. (2002) Global changes in gene expression in response to high light in Arabidopsis. Plant Physiol. 130: 1109-1120.

Seki M., Narusaka M., Abe H., Kasuga M., Yamaguchi-Shinozaki K., Carninci P., Hayashizaki Y., Shinozaki K. (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13: 61-72.

Schenk P.M., Kazan K.,Wilson I., Anderson J.P., Richmond T., Somerville S.C., and Manners J.M. (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. PNAS 97: 11655-11660.

Solomon P.S., Oliver R.P. (2001) The nitrogen content of the tomato leaf apoplast increases during infection by Cladosporium fulvum. Planta 213: 241-249.

Solomon P.S., Oliver R.P. (2002) Evidence that gamma-aminobutyric acid is a major nitrogen source during Cladosporium fulvum infection of tomato. Planta 214: 414-420.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences 98: 5116-5121.

Wagner U., Edwards R., Dixon D.P., Mauch F. (2002) Probing the diversity of the Arabidopsis glutathione S-transferase gene family. Plant Mol Biol. 49: 515-532.

5. Identification of putative *L. maculans* knockout mutants from a T-DNA insertion library

5.1. Introduction

One method for predicting gene function is to draw conclusions about a genes role based on similarity to characterized sequences. However all inferences of this type are based on annotations and gene predictions in the sequence databases and therefore are limited by the number of annotated genes and the quality of annotated data. Similarly, expression analysis can provide valuable data as to timing and strength of gene activity in response to certain stimuli. This information can provide clues as to the function and relative importance of a particular gene, however the ultimate test of gene function is to disrupt expression and analyze the effects on phenotype that result. The disruption of a gene of interest to determine its function is commonly termed reverse genetics. This is opposed to a forward genetic approach, which begins with phenotypic differences and attempts to identify the genes responsible for those differences.

It is possible in many fungi to selectively knock out genes through simple homologous recombination between a native gene and a knockout construct. However the efficiency of gene targeting and length of homologous sequence required in the construct can range from very highly efficient with short homologous regions as in *Saccharomyces cerivisea* and *Schizosaccharomyces pombe,* to very low efficiency with long regions of homology as in *Cladosporium fulvum* (Krawchuk and Wahls 1999; Segers et al. 2001). It has proven possible to perform homologous gene targeting in *L. maculans*, although this required a minimum of 6.9 kb of homologous sequence and the screening of 23 transformants (Wilson et al. 2002). Even with the addition of a negative selection marker to increase the number of transformants recovered that had inserted via recombination; targeting constructs still contained 7.2 kb and 8.5 kb of homologous sequence and required the screening of 6 and 20 transformants respectively to identify a disruption event (Gardiner and Howlett 2004).

When selectively targeting genes for inactivation is not feasible, an alternative approach is to create large numbers of mutants and screen for those individuals lacking an operative copy of your gene of interest. The plant pathogenic bacterium *Agrobacterium tumefaciens*, commonly used to introduce foreign genes in a variety of plant species does so by randomly integrating a small piece of its DNA (T-DNA) into the host genome. Thus, in addition to being an agent for genetic transformation it is also mutagenic. This mutagenic property has been exploited in *A. thaliana* to create libraries of mutant lines (Krysan et al. 1996). These libraries have a number of features which make them an incredibly valuable resource for studying gene function (reviewed by Krysan et al. 1999). First, they are collections of plant lines which can be maintained indefinitely as seed stocks. Second, each mutant is 'tagged' because mutations are caused by the insertion of a DNA fragment of known sequence. Third, if the library is large enough it should contain an insertion in nearly every gene. Finally, these libraries can be used for either forward or reverse genetic screening.

Agrobacterium also has the ability to transform a number of filamentous fungal species including *L. maculans* (de Groot et al. 1998; Gardiner and Howlett, 2004). It is thus possible to use *Agrobacterium* transformation to develop gene knockout libraries in filamentous fungi species. The development of gene knock out libraries is expected to be relatively easy in these species due to the shorter generation times of fungi and the ability to accommodate large numbers of strains in the lab, rather than a greenhouse. Also, fungal genomes have an average size of 36 Mb and gene densities of approximately 220-260 genes/Mb (Kupfer et al. 1997). This smaller size and higher density increases the probability of a random T-DNA insertion occurring within a gene and reduces the size of the library required to achieve saturation.

A library of 2784 *L. maculans* insertional mutants was created to study gene function in this species. The initial focus of this research was to examine potential pathogenesis genes, defined as those required for growth and completion of the life cycle *in planta* but not *in vitro*. Genes with a possible role in pathogenesis were selected from the set of ESTs described in chapter three. A PCR based reverse genetics approach as described by Krysan et al. (1996) was used to search the library for lines containing insertions near these genes of interest. Four lines with insertions near one of the selected genes were identified. One appears to exhibit reduced pathogenicity on *B. napus*. The probable function of this gene and its role in pathogenicity are discussed.

5.2. Materials and methods

5.2.1. Fungal and Bacterial strains

L. maculans isolate WA74, from the collection at the AAFC Saskatoon Research Centre was prepared by culturing surface sterilized, infected Westar cotyledons on V-8 agar plates containing 100 μg/ml streptomycin. Plates were incubated at room temperature under continuous fluorescent light. After 2 weeks, spores were collected in sterile water. Spore suspensions were filtered and the concentration was adjusted to 1x 10^7 spores/ml. Spore suspensions were stored at -20°C. *Agrobacterium* strain A973 containing plasmid pAD1624 (described by Abuodeh et al. 2000) was obtained from the University of Arizona. *Agrobacterium* stocks were stored in 40% glycerol at -80°C.

5.2.2. Transformation of *L. maculans* strain WA74

Pycnidiospores of WA74 were transformed using a modification of the procedure outlined by Abuodeh et al. (2000). A total of $1\text{x}10^7$ spores were mixed with approximately $1\text{x}10^9$ *Agrobacterium* cells to produce a concentrated mixture which was deposited on a 2 cm^2 piece of Hybond N nylon membrane (Amersham Biosciences). The transformation mixture was incubated in the dark at room temperature for three days. After three days cells were dislodged from the filter paper and plated on solid Czapek-Dox media containing 40 μg/ml Hygromycin B (Sigma-Aldrich) to select for fungal transformants and 80 μg/ml Kanamycin (Sigma-Aldrich) to counter select against *Agrobacterium*. Selection plates were incubated at room temperature for 14 days to allow the growth of transformant colonies.

5.2.3. Construction of the transformant library

Individual transformant colonies were picked and re-inoculated on individual plates of Czapek-Dox agar containing 40 μg/ml Hygromycin B (Sigma-Aldrich) and 80 mg/ml Kanamycin (Sigma-Aldrich). Individual transformant plates were incubated under fluorescent light at room temperature for 5-7 days until transformant colonies were sporulating. Spores were collected from each transformant by flooding plates with 5ml of sterile water. Spore suspensions were transferred to 15ml polystyrene centrifuge tubes (Falcon) and centrifuged. Spore pellets were resuspended in 250 µl of sterile water. A 50 µl aliquot of each spore suspension was mixed with 150 µl of 80% glycerol. Spores in glycerol from individual transformants were catalogued and stored in 96 well microtitre plates at -80°C. The remaining 200 µl of transformant spore suspensions were pooled into 96 well blocks, 5 transformants per well, for DNA extraction.

5.2.4. DNA extraction from pooled transformant spores

Pooled spores were pelleted and resuspended in 1 ml sterile water to wash. Spores were centrifuged again and the supernatant was discarded. The 96 well plates were covered with air permeable tape and left in a freeze dryer for two days. Dried spores were ground to a fine powder by the addition of a single 3mm glass bead to each well followed by shaking at a frequency of 30/s for 3 min. in a mixer mill (Retsch). 100 µl guanidine extraction buffer (6M guanidine-HCL, 0.1M soduium acetate, 0.5% N-lauryl sarcosine) was added to the spore powder. After mixing, the extract was diluted with 100 µl TE, RNAse H was added. The spore extract was drawn through diatomaceous earth columns made according to the directions of Hansen et al. (1995) via the application of vacuum pressure. Columns were washed once with 500 µl extraction buffer lacking N-lauryl sarcosine and again with 500 µl TE/ethanol wash buffer described by Hansen et al. (1995). Genomic spore DNA was eluted from the columns in 100uls TE.

5.2.5. Screening the transformant library

Genomic DNA was stored in 96 well boxes, each well representing five transformant. Following the multiplex strategy for identifying insertion lines outlined by Apiroz-Leehan and Fedmann (1997) sets of superpools were made. One pool represented the entire library of transformants, a set of box pools represented all transformants in a given box, a set of row pools represented all transformants in a given row within a box, and a set of column pools represented all transformants within a given column within a box. For initial screening, the DNA superpool representing all transformant strains was used as the template in PCR reactions with combinations of one primer specific to the left or right border of the T-DNA insert and one primer specific to a gene of interest (Table 5-1). All primers were designed according to the guidelines developed for reverse genetic screening by the University of Wisconsin Arabidopsis Knockout facility (www.biotech.wisc.edu/NewServicesAndResearch/Arabidopsis/Guidelines.asp) and had melting temperatures near 65°C (Table 5-2). The PCR amplifications were performed with 40 ng of pooled template, 10 pmol of each primer, 1.5mM $MgCl_2$, and 0.2mM dNTPs with the following cycle: 95°C for 10 min. 40 cycles of 95°C for 50s, 63°C for 50s, 72°C for 5min. followed by a final step at 72°C for 7 min. Primer combinations yielding a PCR product with the entire library pool template were used in PCR reactions under the same conditions with the six box pool templates. These primer combinations were then used in PCR reactions with all the row and column pools for any box that gave a PCR product in the previous reactions. PCR products were observed either as visible bands on agarose gels or as autoradiograph signals on Southern transfers of PCR gels probed with the target ESTs. When a particular well was identified by cross referencing row and column pools yielding a PCR product the five individual transformants represented in that well were retrieved from -80°C storage. A PCR reaction was

Table 5-1. List of ESTs chosen for reverse genetic screening and their closest BLAST search matches.

Candidate EST	BLAST hit annotation	accession number	score	e-value
LM9-194R	putative cyanide hydratase [Leptosphaeria maculans]	(AF192405)	365	1e-100
LM6-5R	glutamine synthetase [Schizosaccharomyces pombe]	(NC_003424)	339	2e-92
LM6-115R	glutamine synthetase [Aspergillus nidulans]	(AF333968)	303	8e-82
LM6-195R	MALATE SYNTHASE, GLYOXYSOMAL Emericella nidulans	pir||S17773	303	3e-94
LM6-220R	multidrug resistance protein 1 [Aspergillus fumigatus]	(U62933)	244	6e-64
LM9-336	probable autophagy protein AUT7 [Neurospora crassa]	(AL670002)	233	2e-60
LM9-257R	Alcohol dehydrogenase I Neurospora crassa	pir||T49440	231	5e-60
LM6-220F	multidrug resistance protein 1 - fungus (Filobasidium floriforme)	(U62929)	187	6e-47
LM6-269R	methionine synthase [Cladosporium fulvum]	(AF226997)	175	1e-47
LM6-291R	methionine synthase [Cladosporium fulvum]	(AF226997)	142	1e-39
LM6-127F	multidrug resistance protein 1 - fungus (Filobasidium floriforme)	(U62929)	135	2e-31
LM6-160F	aromatic amino acid and leucine permease [Penicillium chrysogenum]	(AJ458939)	132	3e-30
LM6-147F	Catalase B [Aspergillus fumigatus]	(Y07763)	99.8	1e-20
LM6-127R	multidrug resistance protein 1 [Aspergillus fumigatus]	(U62933)	72.8	2e-12
LM9-201R	Possible pathogenesis-related protein precursor [Aspergillus fumigatus]	(AL683874)	62.8	2e-09
LM9-205	trypsin-like protease [Phaeosphaeria nodorum]	(AF092435)	316	4e-85
LM6-233	trypsin-like protease [Phaeosphaeria nodorum]	(AF092435)	316	3e-85
LM6-136R	glutamine synthetase [Coccidioides immitis]	(AF306534)	167	5e-41
LM9-361F	argininosuccinate synthetase [Saccharomyces cerevisiae]	(M35237)	91.3	2e-18
ILS-1148	t-complex protein 1, alpha subunit [Schizosaccharomyces pombe]	(NC_003423)	322	2e-87
ILS-772R	translation elongation factor 3 [Ajellomyces capsulatus]	(AF057023)	306	7e-83
ILS-598	Exopolygalacturonase precursor (ExoPG)	(AAC26146.1)	294	5e-79
ILS-662R	cytochrome c oxidase polypeptide II [Hypocrea jecorina]	(NC_003388)	287	6e-77
ILS-656	bic-P1-semicolon- Enhancer of Bicaudal [Drosophila melanogaster]	(NM_057505)	154	1e-36
ILS-217	rhamnogalacturonan acetyl esterase [Aspergillus niger]	(AJ242854)	117	9e-26
ILS-325R	transcription factor [Hypocrea jecorina]	(AJ413272)	112	4e-24
ILS-172	protein synthesis inhibitor, putative [Thermotoga maritima]	(NC_000853)	111	1e-23
ILS-409R	zinc cluster transcription factor Fcr1p [Candida albicans]	(AF057038)	89.7	2e-17
ILS-251F	beta-glucosidase 5 [Coccidioides immitis]	(AY049946)	84.3	6e-16
ILS-1220	glycoprotein CIH1 [Colletotrichum lindemuthianum]	(AJ001441)	75.1	7e-17
ILS-772F	Elongation factor 3 (EF-3) - Pneumocystis carinii	(M87665)	167	5e-41

Table 5-2. Primer sequences used for reverse genetic screening.

Candidate EST	Left Primer sequence	Tm (°C)	Right primer sequence	Tm (°C)
LM9-194R	CGTCCGAAAACAAAGACAATCACCATTTC	63.9	ACGAGCTTTTCAACGTGTGTAGGCTTGAT	65.3
LM6-5R	TTCTGGTTTGGCCTGGAGCAAGAGTACA	66.6	TACATGTCTCGCGATATGCCGGTGATAC	66.6
LM6-115R	TCAAAGTCAAAGACGCTCGACAAGAAGGT	65.3	AGAGATGTTGATCTCGGCGTACATGCAG	66.6
LM6-195R	AAAGCGCTCCTCCAGCGAAGAGTAATTC	66.6	CGTTGTTGAAGAAGTAGAGCCCAAAGTCG	66.7
LM6-220R	ATGATTGTCGTTCTACCCTCGGCTGTTTT	65.3	CCTTCCTCACGGATATGATCAAGAGGTTG	66.7
LM9-336	AATAGCTAGACCCAGGCATTTCCACCAG	66.6	TGAATCTACACCCACATCCACACCCACA	66.6
LM9-257R	AAAACTAAATCGCCTCTGCGCTATCCTTG	65.3	CCCAGATGAAGTCCTCGTCAACATCAAAT	65.3
LM6-269R	CACTCCCACTTCTTGTTTCTCTTCCCATC	66.7	TAGTTGGAGTCGAACCACTTGCCATTTC	65.1
LM6-291R	CACCCACATCCATAATGGTTCACTCTTC	65.1	AGATCTCCATGCCTTTTTAATTCGCCAAC	63.9
LM6-127F	TGAGCTGCCCGAAATCAAAATCAACTACT	63.9	GGTCCCAATTGTACATCAAGTCCGAAAAA	63.9
LM6-160F	TCTTCCTGTCACTGGATCATTTGTTCGTC	65.3	TTCAAAGACTGGAGAATTGGCTCACAGGT	65.3
LM6-147F	CCCAAGCACAATGCCCTTACATGAGTAAT	65.3	AATCGGCTTACCCCACTGGTATCCTTCTT	66.7
LM9-201R	AGCCAGATGTTCCCGACGAAGTTTTCCA	66.6	CCATTGCCATAGATGGGTATGAACAGAGA	65.3
LM9-205	AAACAACCATCACCATGCAGATCAAAGAC	63.9	TCGTCATGAAAGCTTTTTCACAACCCAAT	62.4
LM6-233	AAACAACCATCACCATGCAGATCAAAGAC	63.9	TCGTCATGAAAGCTTTTTCACAACCCAAT	62.4
LM6-136R	CCTTGTGCAAACTCGAGGCTTGTCTACTT	66.7	AACCGTCAAAGTTCCATTCTGGGAGGTC	66.6
LM9-361F	GTCATCTATCCTATGCACGGCTCATGTTG	66.7	CAGTTGTACAATGCATGTCTTCTCGCTGA	65.3
ILS-1148	AGTCAAGTACCCTGTCAAGGCGGTAAACA	66.7	CCACGCAGAATGATAGATGCAGAAGAGAA	65.3
ILS-772R	AGTTCTTGAACAACGTCATTCAGCACGTC	65.3	TTACCCATCGCATCGAAGGTTACTTCATC	65.3
ILS-598	TACCAGCCAGATAAACCCTTCCCTGTTTC	66.7	GAAATTCAAGTCCTCAATACGGCCACCTT	65.3
ILS-662R	AGAATAGGGCCTCATAACCAAGATGTTTT	62.4	GTAGCAATTCTAACCCCATTACCTGTTCA	63.9
ILS-656	ACAACCGCAAATATGGATCAAGCAAAGTT	62.4	CATTTCCCTGCTGTGGTTCTACCAGGATA	66.7
ILS-217	CGGTGGCAAGAACATCACAGTACACACAT	66.7	CAAGTGCGTATAGTTCCACGGAAGAACCT	66.7
ILS-325R	ACCTTCTTCTCGCAGACCATCAATACCAG	66.7	ATGATCCTATCTTTTTGCGCCGTTCAAAT	62.4
ILS-172	GCCGCTATGGCTATCTGGAAGGAATAGAA	66.7	CGGGAGGAAAGGAAGTTGCGTATTGATAA	65.3
ILS-409R	ATTGAGCACATACCACGATCCGCCAAAC	66.6	CGTTCAAGGAGCTTTTGTTGCATTCGAT	63.7
ILS-251F	AGACATCAAGGAGAAGGCTCCCAAGGATA	66.7	TCTTGAAGAAGAAGACGCAGGCCTTGCTA	66.7
ILS-1220	TCATCGCCAGCAACTATACGTCTGGTATT	65.3	ATAGCATGTAATTAAGCTCGGGCGGACCTC	68.1
T-DNA	CATTTTATAATAACGCTGCGGACATCTAC	62.4	TAAGAGCGAATTTGGCCTGTAGACCTCAA	65.3

performed on genomic DNA from each of these individuals under the same conditions as listed above to identify the particular transformant bearing a T-DNA insertion near the gene of interest.

5.2.6. Sequencing reactions and analysis

Sequencing reactions were carried out as described in chapter three. Finished edited sequences were compared to previously studied, publicly available sequences using the BLAST or InterProScan analysis programs (Altschul et al. 1996; The InterPro Consortium 2001).

5.2.7. Southern hybridization

Southern hybridization was performed as described by Sharpe et al. (1995). Transformant genomic DNA (5μg) was digested with either Hind III or Eco RI. Digested DNA was size separated on agarose gels and blotted onto Hybond N nylon membranes (Amersham Biosciences) using a standard alkali transfer protocol. Southern blots of DNA were probed with a 2.7 kb Sac I/Bgl II pAD1624 restriction fragment containing the Hygromycin resistance gene and right border of the T-DNA construct (Figure 5-1).

5.2.8. Cotyledon inoculations

Three cultivars of *B. napus* showing a differential response to a variety of *L. maculans* strains were infected with wildtype and transformant strains of *L. maculans* to assess pathogenicity. Inoculations were performed on plants at the cotyledon growth stage as described in chapter two.

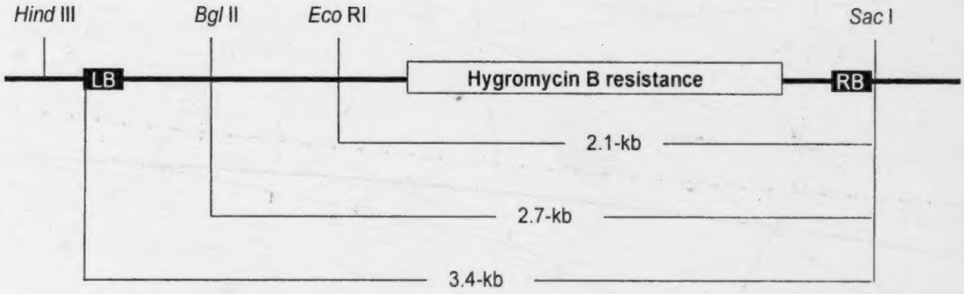

Hind III
Bgl II
Eco RI
Sac I
LB
Hygromycin B resistance
RB
2.1-kb
2.7-kb
3.4-kb

Figure 5-1. The Hygromycin resistance T-DNA construct. The development of the *A. tumefaciens* vector pAD1624 containing this fungal transformation construct was described in Aboudeh et al. 2000. LB, left border sequence. RB, right border sequence. The 2.7 kb *Sac* I/*Bgl* II restriction fragment was used to probe southerns of transformant DNA digested with either *Eco* RI or *Hind* III.

5.3. Results

5.3.1.Construction of the *L. maculans* T-DNA insertion library

A total of 2784 individual *L. maculans* transformant lines were generated. A random sampling of 20 lines from this library were used in Southern hybridizations with a 2.7 kb probe spanning the right border and hygromycin B resistance gene to gain information about the copy number and distribution of insertions (Figure 5-2). Of the 20 lines tested, digestion with Hind III, which does not cleave within the T-DNA construct, indicated that the number of insert loci ranged from one to four with an average of 1.45 insertions per line. Only five of the transformant lines tested exhibited twice as many bands when digested with Eco RI, which cuts within the probe sequence. A further eight of the transformants showed only one band in the Eco RI digest. The remaining seven transformants gave more than twice as many bands in the Eco RI digest, indicating the presence of complex insertion events, which are common features of T-DNA transformation (Kononov et al. 1997; De Buck et al. 1999). Insertions appeared to occur at random loci as all bands observed were of different sizes indicating their presence in different genomic fragments.

5.3.2. Screening the *L. maculans* insertion library

L. maculans has a genome size that ranges from approximately 23 to 34 Mb depending on the isolate (Morales et al. 1993; Cozijnsen et al. 2000). A knockout library of only 2784 individual transformant strains is insufficient to saturate the genome. If every insertion had occurred in a different random 1 kb segment, genome coverage would only

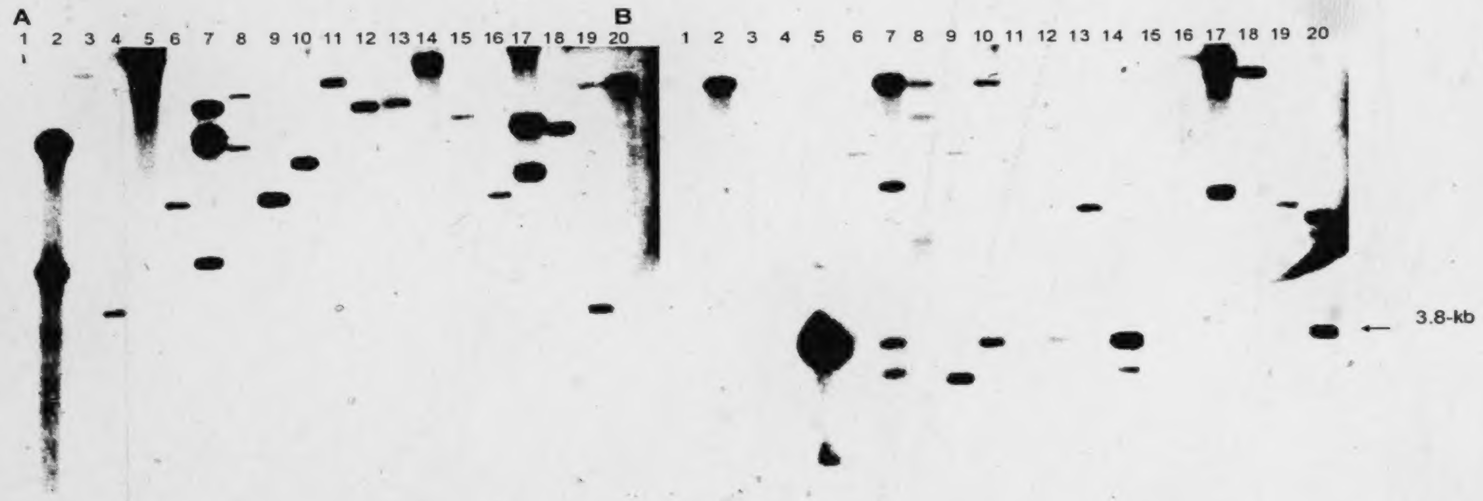

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
3.8-kb

Figure 5-2. Southern analysis of T-DNA insertion events. Genomic DNA from 20 lines of the *L. maculans* insertion library was digested with either *Hind* III (**A**) or *Eco* RI (**B**). Both digest gels were blotted and probed with a 2.7 kb *Sac* I – *Bgl* II fragment of the T-DNA construct containing the hygromycin resistance gene. The recurring 3.8 kb *Eco* RI fragment indicating the likelihood of a tandem T-DNA insertion is marked with the arrow.

reach approximately 10%. Using the equation presented by Krysan et al. (1999) $P = 1 - (1 -[x/y])^n$ where P= probability of finding a T-DNA insert within a given segment, x = length of the given segment in kilobases, y = size of the genome in kilobases and n = the number of T-DNA insertion mutants, it is possible to derive an approximate probability of finding a T-DNA insertion within a gene of interest. For this particular library, using an estimated genome size of 30 Mb the probability of finding a T-DNA insert within a 1 kb target site is 8.86%. Although the probability of finding a T-DNA insert near any one particular gene of interest is low, the probability of finding an insert near one member of a set of genes rises as the set size increases. From the EST library described in chapter three, a set of 28 genes with a potential role in pathogenesis was identified (Table 5-1). The probability of finding an insertion near one of these genes is 92.56%.

The set of 28 potential pathogenicity EST sequences were used to design PCR primers that would amplify their corresponding genes according to the guidelines presented for screening knockout libraries in *A. thaliana* (www.biotech.wisc.edu/NewServicesAndResearch/Arabidopsis/Guidelines.asp). Combinations of one EST specific primer and one T-DNA specific primer were used in reactions with genomic DNA representing the entire knockout library as template to screen for the presence of insertions near these genes of interest. In the initial whole library screen, PCR products were observed with EST specific primers from seven of the candidate genes. Genomic DNA representing each box of pooled transformants was used as template in PCR reactions to browse the library for the individual transformants containing the insertion events identified. Only four of these products could be reproduced using the box pools as template. These products were generated using primers specific to the ESTs LM9-194R, LM9-336, LM6-136R and ILS-409R. The EST LM9-194 sequence is identical to that of the *L. maculans* cyanide hydratase gene, save for a 2-bp mismatch (AF192405.1). The EST LM9-336 has a translated protein sequence very similar to the *idi7* gene of *Podospora anserina* which is involved in autophagy (AAN41258.1; score 231 e-value 2e-59).The EST LM6-136R has a translated protein sequence very similar to the glutamine synthase gene of *Coccidioides posadasii* (AAG28781.1; score 167 e-value 8e-41). Finally, the EST ILS-409R has a translated protein sequence similar to the *Fcr1* gene of *Candida albicans* (093870; score 90, e-value 4e-17).The three primer combinations not reproduced in the box pool reactions were retested on the complete library template, and gave inconsistent results. Screening did not proceed any further with these putative insertion events. The four insertion events yielding reproducible PCR products were localized to particular wells in the appropriate box of pooled transformant DNA by cross referencing row and column pools. Following the identification of a particular well containing DNA with the T-DNA fragment inserted near a candidate gene, the five individual transformant strains represented in the well pool were retrieved from storage, and cultured for DNA extraction. A final round of PCR reactions using individual transformant strain DNA as template identified the particular transformant lines containing these T-DNA inserts near genes of interest. Southern transfers indicated that LMT-98 (insertion near LM6-136), LMT-1851 (insertion near LM9-336), and LMT-2162 (insertion near ILS-409) each contained a single T-DNA insertion locus. The remaining transformant line, LMT-1240 (insertion near LM9-194) appeared to contain two T-DNA insert loci.

5.3.3. Sequencing T-DNA insertion sites

For each of the four mutant strains identified from the reverse genetic screening of the *L. maculans* transformant library, the insertion specific PCR products observed from the screening were isolated and sequenced. Wildtype genomic sequences were also obtained from PCR products using the EST specific primers designed for screening the transformant library (Figure 5-3). A comparison of the EST sequence for clone ILS-409R and its genomic PCR product reveals the presence of two small introns of 41 and 54-bp respectively. The EST sequence for ILS-409R has significant similarity to hypothetical proteins from *Aspergillus nidulans* and *Neurospora crassa* as well as to the *Candida albicans Fcr1p* zinc finger transcription factor (score 90, e-value 4e-17 Genbank accession number O93870) mediating pleiotropic drug resistance and to a lesser extent the *Cat8* transcription factor involved in gluconeogenesis (score 68, e-value 1e-10 Genbank accession number AAF35321). The sequence read from the right border specific T-DNA primer has no significant similarity to any known sequences at either the nucleotide or protein levels. The PCR product obtained from the right EST specific primer and the right border specific T-DNA primer is approximately 1.7 kb. Sequence data is available for the 435-bp extending from the right T-DNA border and 669-bp extending off the right EST specific primer leaving an internal sequencing gap of approximately 600-bp in this PCR product. The R in the ILS-409R EST clone name indicates that this EST is derived from half of the original cDNA library insert and corresponds to the 5' end of this particular gene. The cDNAs used to construct the libraries in chapter two were size selected to enrich for full or near full length sequences. If it can be assumed that the EST sequence extends close to the 5' end of the gene, the T-

DNA must have then inserted nearly 1 kb upstream of this particular gene in the *L. maculans* genome (Figure 5-3).

The primers designed from the LM6-136R EST were separated by 352-bp of EST sequence. When used on genomic DNA these same primers produce an amplified DNA fragment approximately 1 kb in size. Sequence is available for 579-bp of this genomic PCR fragment extending from the left EST specific primer and for 201-bp of this genomic PCR product extending out from the right EST specific primer. The Left primer sequence matches the

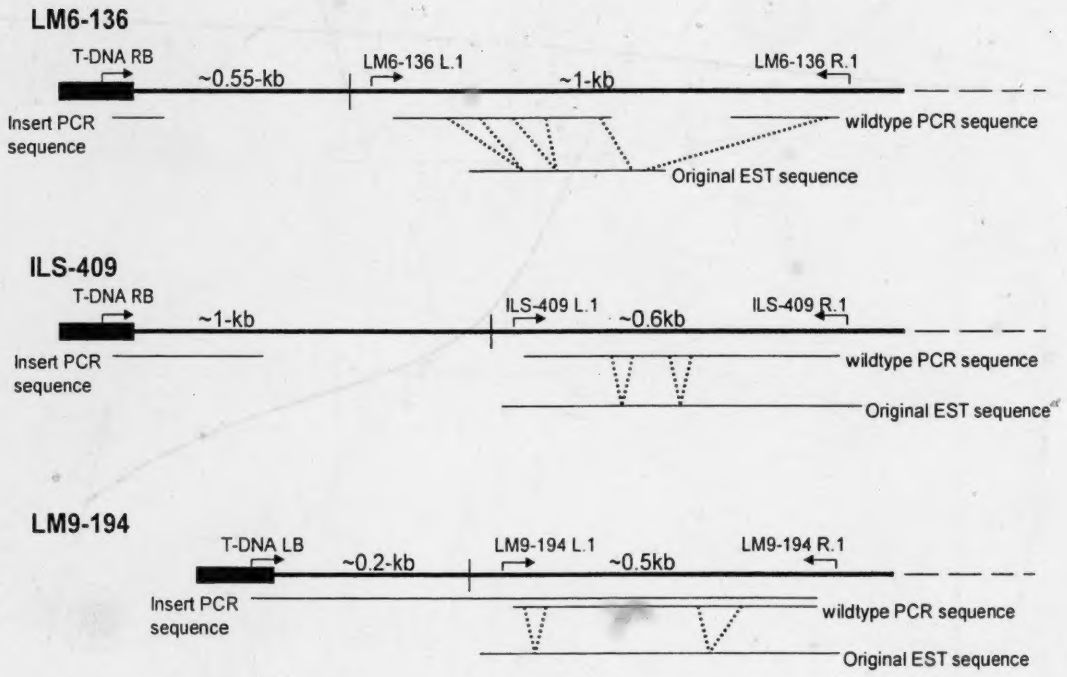

LM6-136
T-DNA RB
~0.55-kb
LM6-136 L.1
~1-kb
LM6-136 R.1
Insert PCR sequence
wildtype PCR sequence
Original EST sequence
ILS-409
T-DNA RB
~1-kb
ILS-409 L.1
~0.6kb
ILS-409 R.1
Insert PCR sequence
wildtype PCR sequence
Original EST sequence
LM9-194
T-DNA LB
~0.2-kb
LM9-194 L.1
~0.5kb
LM9-194 R.1
Insert PCR sequence
wildtype PCR sequence
Original EST sequence

Figure 5-3. Hypothetical insertion structure and available sequence data. Heavy lines represent the hypothetical insertion sites for each of the T-DNA constructs identified in the reverse genetic PCR screen. The vertical lines represent the hypothetical start codon for each gene. Bent arrows represent the binding sites for the primers used for screening and sequencing. The blocks represent the T-DNA sequence ends. Narrow lines beneath each hypothetical insertion map indicate areas near the T-DNA insertion sites for which sequence data is available. Dashed lines indicate the approximate position and size of introns as determined by comparison between the original EST sequences and the sequence of genomic wildtype PCR products.

EST sequence and reveals the presence of three intronic regions of 73 and 76-bp plus one of approximately 500-bp, encompassing the entire genomic PCR sequence from the right EST specific primer. The PCR product generated by the right border T-DNA primer and the right EST specific primer is approximately 1.5 kb, indicating that the T-DNA insertion is ~500-bp upstream of this gene (Figure 5-3). Sequence of the knockout PCR product extending from the T-DNA right border has no significant similarity to anything in the Genbank database at either the nucleotide or predicted protein sequence levels.

The LM9-336 EST specific primers amplify a 729-bp sequence when used in a reaction with wildtype genomic DNA as template. Comparison of this sequence with the EST sequence shows the presence of two introns, 57 and 60-bp. The knockout PCR product generated from amplification with the right EST specific primer and the right border specific T-DNA primer is 551-bp and is sequenced in its entirety. Based on the size of the wildtype and knockout PCR products it would be expected that the T-DNA insertion had occurred within the coding sequence of this gene. However, there is absolutely no similarity between the wildtype PCR product and putative knockout PCR product sequences. Further, the knockout sequence has no appreciable similarity to any of the sequences present in the Genbank database.

Finally, the wildtype and knockout PCR products for LM9-194R differ in size by 231-bp. The knockout PCR product was generated through amplification with the right border T-DNA product and the right EST specific primer. Both PCR products have been completely sequenced and align perfectly with each other and with the original EST sequence, except for two introns of 48 and 92-bp missing from the EST sequence. In this case it is absolutely clear that the T-DNA sequence has inserted 149-bp upstream of the

EST coding sequence (Figure 5-3).

5.3.4. Phenotypic testing of putative knockouts

Each of the four putative knockout strains of *L. maculans* identified was tested for any effects the T-DNA insertion may have had on their pathogenicity by inoculating *B. napus* cotyledons (Table 5-3). Testing was performed on 14 to 22 individuals each of the three differential *B. napus* cultivars Westar, Quinta and Glacier. These cultivars are commonly used to distinguish between *L. maculans* pathogenicity groups 2, 3 and 4. The wildtype *L. maculans* strain used for transformation, WA74, belongs to pathogenicity group 2, meaning that it is virulent on Westar, but avirulent on Glacier and Quinta. The transformant LMT-1851, which has a T-DNA insertion identified by the LM9-336 EST specific primers, appears to have slightly elevated pathogenicity on both Westar and Quinta when compared to the WA74 wildtype control. Similarly, the transformant LMT-98, possessing an insertion near the gene corresponding to the LM6-136R EST appears to have slightly reduced pathogenicity on Quinta and Glacier. Cotyledons with ratings between 0 and 3 are considered to be highly resistant, between 4 and 6 are intermediate or moderately resistant and between 7 and 9 are considered to be highly susceptible. The observed increases or decreases in the pathogenicity of these two transformants would not change their classification as either virulent or non-virulent. In contrast, the transformant LMT-2162, possessing a T-DNA insertion upstream of the sequence represented by the ILS-409R EST has virtually no infectivity. Westar, a cultivar normally susceptible to *L. maculans* isolate WA74 is completely resistant to LMT-2162 with an average disease rating of 0.6 ± 0.83. The only visible response to infection was a slight

Table 5-3. Ratings of disease severity caused by putative knockouts on differential *Brassica napus* cultivars.

		Brassica napus cultivar		
Fungal isolate	EST sequence	Westar	Quinta	Glacier
		rating	rating	rating
WA74		7.06 ± 1.68	1.86 ± 1.49	5.06 ± 1.44
LMT-98	LM6-136R	7 ± 1.48	0.91 ± 0.87	3 ± 1.41
LMT-2162	ILS-409R	0.6 ± 0.83	0.69 ± 0.87	0.6 ± 0.51
LMT-1851	LM9-336	8.14 ± 1.39	2.95 ± 1.47	5.11 ± 1.56
LMT-1240	LM9-194R	6.92 ± 1.38	1.95 ± 0.97	4.79 ± 1.32

Ratings represent the average disease ratings for all plants tested in each strain/cultivar combination plus or minus one standard deviation.

darkening around the cotyledon wound site indicating a host response in half of the plants tested. In the other half of the Westar cotyledons inoculated with LMT-2162 there were no observable host reactions or disease symptoms. LMT-2162 exhibited normal growth and sporulation when cultured on complete medium; however this transformant strain had completely lost its ability to successfully infect the *B. napus* cultivar Westar.

5.3.5. Analysis of ILS-409R sequence

The EST nucleotide sequence of ILS-409R has no significant matches in the Genbank nr database, the closest matches belonging to two 21-bp sequences found in nine various *Homo sapiens* and *Drosphila melanogaster* submissions (each with score 42, e-value 1.5). This ILS-409R nucleotide sequence does show some similarity to eight different *Aspergillus nidulans* ESTs in the Genbank EST database (each with score 64 and e-value 4e-07). A region of the protein translation of the ILS-409R EST sequence spanning bases 174 to 380 has high similarity to both hypothetical proteins identified from *Neurospora crassa* , *Magnaporthe grisea* and *A. nidulans* genomic sequences as well as 22 different Zn(2)-Cys(6) type transcription factors. Most similar among these transcription factor matches are the *Candia albicans* Fcr1 fluconazole resistance gene (093870 score 90, e-value 4e-17) and the *C. albicans Cat8* transcription factor involved in gluconeogenesis (AAF35321.1 score 68, e-value 1e-10) (Figure 5-4). This region of the ILS-409R sequence that exhibits similarity to the above mentioned transcription factors contains a translated hypothetical protein sequence conforming to the Zn(2)-Cys(6) motif $CX_2CX_6CX_{5\text{-}9}CX_2CX_6C$ found in a wide variety of fungi (Schjlering and Holmberg 1996). Analysis of this sequence with the Interpro scan tool

```
ILS-409R  1 ffplstyhdp pnpahpns.l altppftspg vg.peplshh 40
                                           1 ms
Fcr1

41  ptmsphrtds mdtqdniRKR VCKACDRCRL KKSKCDGASP CSRCKADNAI 90
                      +KR V KACD CR+ KK+KCDG  P C+RC  DN I
3   ddhsihshtq glhhikkKKR VGKACDSCRI KKTKCDGKKP CNRCTLDNKI 52

91  CVFGERKKSQ DKVYPKGYVE MLEQQQTQLV AGLRELYsrl qkgeswpgqp 140
    CVF E+KK++ +K +P GYVE +LE +   L    L +L
53  CVFTEKKKTK EKKHPSGYVE LLEARLDILT RSLEKLiels rphlqfiddi 102

141 lretsgghpl thdilerldl lhpstesskq qaittdskti aiecnkssln 190

103 iteeksvdqe ksspasstpn ssssdhhddv eeqnstgvva pinkvvsyli 152

191 agtlyttagi rqlgfrarsh ffaaiayrdp yhk 223

153 keqgllknip leweqgteia anfdpnrnlk sss 185
```

Figure 5-4. Alignment between the translated *L. maculans* gene represented by the ILS-409 EST and the amino acid sequence of the *C. albicans* Fcr1 (Fluconazole resistance protein 1) gene. The region of alignment is denoted by Capital letters. The shaded area illustrates the conserved $CX_2CX_6CX_{5\text{-}9}CX_2CX_6C$ motif characteristic of Zn(2)-Cys(6) transcription factors. The periods at positions 19 and 33 in the ILS-409R sequence represent stop codons.

(http://www.ebi.ac.uk/InterProScan/) could not identify any other motifs in this sequence.

5.4. Discussion

In order to study gene function in *L. maculans*, a library of randomly mutagenized fungal strains was created by *A. tumefaciens* mediated transformation with a T-DNA construct carrying a Hygromycin B resistance gene. This approach has a number of benefits. First, as T-DNA mutagenesis occurs via the integration of foreign DNA of known sequence into the host genome, mutants are 'tagged'. Secondly, although labor intensive to establish the library, it is relatively easy to perform random mutagenesis via T-DNA transformation in contrast to generating targeted disruptions of genes in this species. Third, having established a durable library of transformant strains for which DNA is available for reverse genetic screening, any number of genes relating to any aspect of *L. maculans* biology may be studied with this resource. Finally, this library can also be used for forward genetic screening where any identifiable phenotypes can be traced back to the genes responsible through the T-DNA tags.

The library described here was used for reverse genetic screening to identify mutations in or near genes that may have a role in pathogenesis. Reverse genetic screening was made possible by the availability of the *L. maculans* EST sequences described in chapter three. Although the EST library was constructed from the PG2 *L. maculans* strain PL86-12, the knockout library was constructed in the PG2 *L. maculans* strain WA74. The decision to switch the strain used for insertion library construction was based on the ease of transformation; WA74 yields more transformant colonies from a single *A. tumefaciens* co-incubation. This was not anticipated to cause any problems as *L. maculans* gene sequences appear to be very well conserved as evidenced by the exact or almost exact sequence matches between western Canadian isolate PL86-12 and Australian isolate M1 for both the cyanide hydratase EST (LM9-194 R), and opsin ESTs (LM9-44) and their cloned counterparts (Genbank accession numbers AF192405.1; AF290180). Indeed, for the four genes studied here, the EST sequences derived from PL86-12 are identical to the corresponding genomic sequences from isolate WA74.

Although based on a small sample size (20), initial testing of the transformant lines gave no evidence to suggest that T-DNA insertion was not occurring at random loci (Figure 5-2). Integration of T-DNA has been observed to occur at random loci in a number of different filamentous fungi (de Groot et al. 1998). Five of the transformants tested exhibited patterns suggesting the presence of simple, single copy insertion events. Seven of the transformants gave complex banding patterns with higher than expected numbers of bands after digestion with the internal probe sequence cutter *Eco* RI. Six of the transformants contained the same band approximately 3.8 kb in size when digested with *Eco* RI. This band would be expected if tandem T-DNA insertions occurred at the same locus. Tandem insertions of T-DNA sequences are observed at frequencies greater than 50% in different plant T-DNA insertion mutant libraries (Rios et al. 2002; Scholte et al. 2002). The final eight transformants had fewer than the expected number of bands when digested with *Eco* RI. Integration of T-DNA in host genomes often occurs with sequence modifications to either the host genome or the T-DNA fragment, including small insertions of 'filler' DNA, small genomic deletions, inclusion of vector backbone sequences or loss of border regions (Kim et al, 2003; Kumar and Fladung 2002; DeBuck et al. 1999). Typically in plant insertions left hand border sequences are truncated 3 to 100-bp, although truncations of up to 1.5 kb have been reported (Tinland 1996). It is possible that the apparent loss of bands expected in the *Eco* RI digest may be due to loss of T-DNA sequences or other imprecision in T-DNA insertion.

Because the *L. maculans* insertion library described here is relatively small, the probability of finding an insert in a particular gene of interest is conservatively estimated at 8.86%. In order to increase the probability of finding putative knockout genes in this collection and prove the utility of such a resource, a set of candidate genes was selected. This set of 28 candidate genes was chosen based on BLAST results from the EST sequences described in chapter three and included were genes that could be assigned a probable function and for which that probable function was conceivably related either directly or indirectly to pathogenesis or growth within a host environment. Insertions near four members of this candidate gene set could reliably be identified. In contrast to *A. thaliana* insertion libraries, PCR products were directly visible on agarose gels, although Southern transfers and hybridizations with EST fragments were performed on the first round of screening gels to confirm the presence of EST sequence. Initially seven putative knockouts were identified, but only four of these could be reliably reproduced in subsequent rounds of screening. It is probable that the three PCR bands initially observed were simply artifacts.

Partial sequences including T-DNA junctions are available for the genomic wildtype and insertion copies of the genes represented by LM6-136 and ILS-409 (Figure 5-3). The complete sequence of the genomic wildtype and insertion copies of the gene represented by LM9-194 is also available (Figure 5-3). Comparisons were made between original EST sequences, their corresponding genomic sequences and sequences intervening between these genes and the inserted T-DNA fragment. Intronic regions in the genomic sequences were identified. Three of the genes examined here (ILS-409, LM9-336 and LM9-194) had two identifiable introns each, ranging in size from 41 to 76-bp, this is consistent with other filamentous fungi like *N. crassa*, where genome sequencing has revealed an average of 1.7 introns per gene and an average intron size of 134-bp, and *A. fumigatus* which has an average of two introns per gene (Galagan et al. 2003; Pain et al. 2004). One *L. maculans* sequence (LM6-136) had three introns, one of which appeared over 500-bp in size, although since the wildtype genomic PCR product was not sequenced in its entirety, it is impossible to be precise about this.

The insertion PCR product generated from the right border T-DNA and LM9-336 right gene specific primers was consistently observed as an approximately 600-bp band in each of the screening gels. The initial screening Southern transfer also identified this band when probed with a mixture of PCR generated fragments of all 28 candidate genes, although the autoradiograph signal was faint. Despite being generated with the T-DNA right border primer and a 28-bp LM9-336 specific primer, the insert PCR product sequence was completely different from the original EST or genomic sequences generated from both LM9-336 specific primers. Neither does this insertion PCR product share any sequence similarity with any of the other 28 candidate genes for which PCR primers were designed. It is possible that a sequence similar enough to the LM9-336 specific primer to allow amplification exists elsewhere in the *L. maculans* genome. Another possibility would be that the T-DNA insertion caused a translocation or some other chromosomal rearrangement. Translocations and chromosomal rearrangements including deletions, duplications and inversions have been regularly observed in T-DNA transformed plants (Castle et al. 1993; Takano 1997; Nacry et al. 1998; Laufs et al. 1999; Tax and Vernon 2001; Forsbach et al. 2003; Laflueriel et al. 2004). Unfortunately the lack of sequence and mapping data make it impossible to determine the exact cause or nature of this abnormal sequence result.

In order to determine whether any of the mutants identified had any detectable change in pathogenicity phenotype, plants of the three differential genotypes Westar, Quinta and Glacier were subjected to infection at the cotyledon growth stage. Only one of the identified lines, LMT-2162 gave results that were qualitatively different from wildtype WA74. In this case the transformant line grew and sporulated normally in complete V8 medium yet was completely unable to infect the *B. napus* cultivar Westar, which is susceptible to wildtype WA74. This transformant strain contains a simple single locus, single copy T-DNA insertion, approximately 1 kb upstream of the ILS-409 sequence. It would be necessary to perform northern blots or RT-PCR to confirm that transcription of the gene corresponding to the ILS-409 EST sequence was either decreased or absent. Crosses involving the putative knockout LMT-2162 strain and a wildtype strain should also be performed to determine whether the loss of pathogenicity phenotype co-segregates with the T-DNA insertion. Retransformation of the putative knockout strain with a wildtype genomic copy of the ILS-409 gene could confirm that the loss of pathogenicity is caused by the loss of this gene function. However, at this stage a very clear phenotype is observable in this strain and is most likely due to the effects of the insertion upstream of the ILS-409 sequence. This sequence contains the conserved Zn(2) Cys(6) zinc finger DNA binding motif common to many transcription factors, and is similar to the Cat8 gene involved in gluconeogenesis and the Fcr1p gene involved in pleiotropic drug resistance from *A. nidulans* and *C. albicans* respectively (Genbank accession numbers AAF35321 and O93870). The degree of similarity to either of these two gene products is relatively limited and only extends across a 200 to 300-bp region covering the zinc finger motif, making it risky to attempt to predict which particular aspect of fungal biology this putative transcription factor mediates. It will be necessary to perform a more detailed characterization of any and all phenotypic changes observable in this strain to ascertain the role of the gene represented by the ILS-409 EST in *L. maculans* growth and pathogenicity. Studies of this putative transcription factor may be aided by the microarrays described in chapter four. Comparison of wildtype and LMT-2162 transcription under various growth conditions may allow the identification of *L. maculans* genes whose transcription is under the influence of the gene corresponding to ILS-409. Probable functions for these genes could provide clues as to the precise function of the gene represented by ILS-409.

In this chapter, the construction of a library of *L. maculans* strains randomly mutagenized by the insertion of a foreign 'tag' DNA was described. This library is suitable for both forward and reverse analysis of gene function in this organism. This library is a lasting resource that can be applied to the study of most aspects of fungal growth and development, excepting those necessary for sporulation or for which loss is immediately and directly lethal. The utility of this library has been demonstrated by the use of a reverse genetics approach to identify insertional mutants near at least three genes having a potential role in pathogenicity with a fourth in question. One of the insertion lines identified has a clearly observable loss of pathogenicity phenotype and presents an opportunity to further study the molecular mechanisms of *L. maculans* infection of *Brassica* species.

Works Cited

Abuodeh R.O., Orbach M.J., Mandel M.A., Das A., Galgiani J.N. Genetic transformation of Coccidioides immitis facilitated by *Agrobacterium tumefaciens*. 2000. J Infect Dis. 181: 2106-2110.

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., and Lipman D.J. Gapped blast and psi-Blast: a new generation of protein database search programs. Nucleic Acids Res, 1997. 25: 3389-3402.

Castle L.A., Errampalli D., Atherton T.L., Franzmann L.H., Yoon E.S., Meinke D.W. Genetic and molecular characterization of embryonic mutants identified following seed transformation in Arabidopsis. 1993. Mol Gen Genet. 241: 504-514.

Cozijnsen A.J., Popa K.M., Purwantara A., Rolls B.D. and Howlett B.J. Genome analysis of the plant pathogenic ascomycete Leptosphaeria maculans; mapping mating type and host specificity loci. 2000. Mol. Plant Pathol. 1:293-302.

De Buck S., Jacobs A., Van Montagu M., Depicker A. The DNA sequences of T-DNA junctions suggest that complex T-DNA loci are formed by a recombination process resembling T-DNA integration. 1999. Plant J. 20: 295-304.

de Groot M.J., Bundock P., Hooykaas P.J., Beijersbergen A.G. Agrobacterium tumefaciens-mediated transformation of filamentous fungi. 1998. Nat Biotechnol. 16: 839-842.

Forsbach A., Schubert D., Lechtenberg B., Gils M., Schmidt R. A comprehensive characterization of single-copy T-DNA insertions in the Arabidopsis thaliana genome.2003. Plant Mol Biol. 52: 161-176.

Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D., Jaffe D., FitzHugh W., Ma L.J., Smirnov S., Purcell S., Rehman B., Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M., Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A., Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D., Roy A., Foley K., Naylor J., Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S., Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A., DeSouza C.P., Glass L., Orbach M.J., Berglund J.A., Voelker R., Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R., Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M., Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B. The genome sequence of the filamentous fungus Neurospora crassa. 2003. Nature.422: 859-868.

Gardiner DM, Howlett BJ. Negative selection using thymidine kinase increases the efficiency of recovery of transformants with targeted genes in the filamentous fungus Leptosphaeria maculans. 2004. Curr Genet. 45: 249-255.

Hansen N.J., Kristensen P., Lykke J., Mortensen K.K., Clark B.F. A fast, economical and efficient method for DNA purification by use of a homemade bead column. 1995. Biochem Mol Biol Int. 35: 461-465.

Kononov M.E., Bassuner B., Gelvin S.B. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. 1997. Plant J. 11: 945-957.

Krawchuk M.D., Wahls W.P. High-efficiency gene targeting in Schizosaccharomyces pombe using a modular, PCR-based approach with long tracts of flanking homology. 1999. Yeast 15: 1419-1427.

Krysan P.J., Young J.C., Tax F., Sussman M.R. Identification of transferred DNA insertions within Arabidopsis genes involved in signal transduction and ion transport. 1996. Proc. Natl. Acad. Sci. USA 93: 8145-8150.

Krysan P.J., Young J.C., Sussman M.R. T-DNA as an insertional mutagen in Arabidopsis. 1999. Plant Cell 11: 2283-2290.

Kupfer D.M., Reece C.A., Clifton S.W., Roe B.A., and Prade R.A., Multicellular ascomycetous fungal genomes contain more than 8000 genes. 1997. Fungal Genet Biol. 21: 364-372.

Lafleuriel J., Degroote F., Depeiges A., Picard G. A reciprocal translocation, induced by a canonical integration of a single T-DNA, interrupts the HMG-I/Y Arabidopsis thaliana gene. 2004. Plant Physiol Biochem. 42: 171-179.

Laufs P., Autran D., Traas J. A chromosomal paracentric inversion associated with T-DNA integration in Arabidopsis. 1999. Plant J. 18: 131-139.

Morales V.M., Séguin-Swartz G., Taylor J.L. Chromosome size polymorphism in *Leptosphaeria maculans*. 1993. Phytopathology 83: 503-509.

Nacry P., Camilleri C., Courtial B., Caboche M., Bouchez D. Major chromosomal rearrangements induced by T-DNA transformation in Arabidopsis. 1998. Genetics 149: 641-650.

Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M., Fosker N., Fraser A., Harris D., Larke N., Murphy L., Humphray S., O'Neil S., Pertea M., Price C., Rabbinowitsch E., Rajandream M.A., Salzberg S., Saunders D., Seeger K., Sharp S., Warren T., Denning D.W., Barrell B., Hall N. Insight into the genome of Aspergillus fumigatus: analysis of a 922 kb region encompassing the nitrate assimilation gene cluster. 2004. Fungal Genet Biol. 41: 443-453.

Schjerling P., Holmberg S. Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. 1996. Nucleic Acids Res. 24: 4599-4607.

Sexton A.C., Howlett B.J. Characterisation of a cyanide hydratase gene in the phytopathogenic fungus Leptosphaeria maculans. 2000. Mol. Gen. Genet. 263: 463-470.

Segers G., Bradshaw N., Archer D., Blissett K., Oliver R.P. Alcohol oxidase is a novel pathogenicity factor for Cladosporium fulvum, but aldehyde dehydrogenase is dispensable. 2001. Mol Plant Microb Interact. 14: 367-377.

Sharpe, A.G., Parkin, I.A.P., D.J. Keith, and D.J. Lydiate. 1995. Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome **38**: 1112-1121.

Takano M., Egawa H., Ikeda J.E., Wakasa K. The structures of integration sites in transgenic rice. 1997. Plant J. 11: 353-361.

Tax F.E., Vernon D.M. T-DNA-associated duplication/translocations in Arabidopsis. Implications for mutant analysis and functional genomics. 2001. Plant Physiol. 126: 1527-1538.

The InterPro Consortium (Apweiler R., Attwood T.K., Bairoch A., Bateman A., Birney E., Biswas, Bucher P., Cerutti L., Corpet F., Croning M.D.R., Durbin R., Falquet L., Fleischmann W., Gouzy J., Hermjakob H., Hulo N., Jonassen I., Kahn D., Kanapin A., Karavidopoulou Y., Lopez R., Marx B., Mulder N.J., Oinn T.M., Pagni M., Servant F., Sigrist C.J.A., Zdobnov E.M.). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 2001. 29: 37-40.

Wilson L.M., Idnurm A., Howlett B.J. Characterization of a gene (*spl*) encoding a secreted protein from *Leptosphaeria maculans*, the blackleg pathogen of *Brassica napus*. 2002. Mol Plant Pathol 3: 487-493.

6. Conclusion and future prospects

This thesis describes the development and application of genomic scale resources to the study of *Leptosphaeria maculans* pathogenicity and *Brassica* defense responses. Investigations of *Brassica* resistance to blackleg disease caused by *L. maculans* have noted B genome sources of resistance as being particularly strong (reviewed in Rimmer and van den Berg 1992). A natural *Brassica juncea* (AABB) population segregating for resistance to *L. maculans* was developed. This population made possible the mapping of the first B genome *L. maculans* resistance gene, *LMJR1*. The use of natural segregating populations rather than interspecific hybrids carrying large, non-recombinogenic introgressions will make possible the proper identification of closely linked genetic markers which can be employed to positionally clone the gene responsible for resistance as well as to breed this trait into susceptible species and varieties more efficiently. Currently, genetic markers flanking *LMJR1* have been positioned 8.7 and 22.1 cM away (chapter two). Unfortunately these distances are too great for these markers to be of use for positional cloning. This population also contains a second, recessive resistance trait. A possible location for this trait was posited, although it could not be mapped conclusively in the populations described in chapter one. Future research efforts should focus on finding molecular markers more closely linked to *LMJR1* to enable cloning this gene. Effort should also be devoted to confirming a map position for the second recessive blackleg resistance genes in this population. This could be readily achieved by taking advantage of F3 families from the AC Vulcan x UM3132 cross which are segregating only for this resistance trait.

Libraries of cDNAs were developed from both *in vitro* cultured *L. maculans* and infected *Brassica napus* leaf tissue. These cDNAs were partially sequenced to generate a collection of ESTs (chapter three). The infected leaf library was enriched for infection-specific sequences and approximately 10% of genes in this collection were from *L. maculans* growing *in planta.* A number of sequences were present which could be associated with either pathogenicity or defense based on nucleotide sequence similarity to previously studied genes. This EST collection provided the basis for further analysis of gene expression in *B. napus* and *L. maculans* and for the study of gene function in *L. maculans*. A total 741 ESTs from *in vitro* grown *L. maculans* and 1281 ESTs from infected *B. napus* leaf tissue were collected. The observed levels of sequence redundancy in these EST collections was low, indicating that further sequencing of the cDNA libraries could rapidly expand these EST collections. This would be of particular use for studying *L. maculans* which is poorly represented in sequence databases **(http://www.ncbi.nlm.nih.gov/;** Howlett et al. 2001).

Those cDNAs which were partially sequenced to generate ESTs were used for the construction of glass microarray slides to study gene expression. Despite a lack of reproducibility between replicate hybridizations using these slides, a set of 76 genes were determined to have significant changes in their expression at various points over the course of an eight day infection. A large proportion (36.8%) of these genes had no previously characterized function, making them valuable targets for further study as some of them may be novel defense or pathogenicity genes. Although a number of observed changes in expression were consistent with previous reports of expression changes for specific genes, some changes such as the apparent downregulation of PR genes were inconsistent with other reported observations. It would be of great value to confirm these results using other techniques for examining gene expression. These microarray slides can be used in other experiments; it would be interesting to make comparisons between infections involving virulent and avirulent fungal strains as well as resistant and susceptible hosts.

Finally a library of *L. maculans* transformants, each carrying a random T-DNA insertion was created as a resource to study gene function in this species (chapter five). Integration of T-DNA was demonstrated to occur at 1.45 loci per transformation on average and appeared random. This library is suitable for both forward and genetic screens, and its utility was demonstrated by performing reverse genetic screening to identify insertions near potential pathogenicity genes. Four putative knockouts were detected and one of these, near the gene represented by the EST ILS-409R, a predicted zinc finger transcription factor, was shown to have lost the ability to infect a susceptible *B. napus* cultivar. This putative knockout will require further characterization to conclusively show that the loss of pathogenicity phenotype is caused by the T-DNA insertion disrupting the function of the ILS-409 gene. As the insertion in this transformant strain occurred upstream of the gene sequence, northern blots or RT-PCR should be performed to resolve whether or not expression of this transcription factor is actually eliminated. Also, crosses involving this transformant and a wildtype strain of *L. maculans* should be made to track whether or not the loss of pathogenicity phenotype co-segregates with the T-DNA insertion. Alternately the wildtype genomic copy of this transcription factor could be cloned and transformed into the putative knockout strain to determine if the pathogenicity phenotype can be rescued. It may also be interesting to compare gene expression in this putative knockout strain during infection to that of a wildtype strain using the microarrays described in chapter four. In this manner genes that are regulated by this putative transcription factor may be identified.

Although the probability of finding a T-DNA insertion in any given 1 kb segment of the *L. maculans* genome is only 8.86% in this library currently, it would be a straightforward procedure to generate more transformants and increase the likelihood of an insertion event near any particular sequence of interest.

Four different complementary resources for studying *L. maculans* infection of *B. napus* have been developed; a collection of *L. maculans* and *B. napus* ESTs including a set enriched for infection specific sequences, a *B. juncea* mapping population segregating for resistance to blackleg, a microarray platform for simultaneously examining gene expression in both *L. maculans* and *B. napus* and a library of random tagged mutant *L. maculans* strains for the study of gene function in this species. These resources may be used in combinations. The *B. juncea* population described in chapter two or other *Brassica* mapping populations can be used to discover the genomic locations of any ESTs of interest. The microarray slides can yield expression information for all the ESTs described in chapter three. These EST sequences provide information for performing reverse genetic screens in the knockout library described in chapter five.

These various genomics resources have proven their value by revealing the genomic location of *LMJR1* a B genome trait conferring resistance to blackleg, tripling the amount of sequence data available for *L. maculans*, identifying previously uncharacterized sequences with changes in expression at various times during infection, and by creating a strain of *L. maculans* that has lost its pathogenicity as the result of a likely disruption in an uncharacterized transcription factor. Each of these different resources has the potential to be expanded. These tools will also greatly enhance the ability of the research community to understand the molecular genetics of the *L. maculans-Brassica* interaction.

Works cited

Howlett B.J., Idnurm A., Pedras M.S. (2001) Leptosphaeria maculans, the causal agent of blackleg disease of Brassicas. Fungal Genet Biol. 33: 1-14.

Rimmer, S.R., and van den Berg, C.G.J. (1992). Resistance of oilseed *Brassica* spp. To blackleg caused by *Leptosphaeria maculans*. Can J Plant Path **14**: 56-66.